CLARIFICATION OF THE ROLE AND CONFIGURATION OF LEARNING AS THEY ARE MANIFEST IN PERFORMANCE AT HALSTEAD'S CATEGORY TEST

By

Kenny William Bertram

A DISSERTATION

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

DOCTOR OF PHILOSOPHY

Department of Psychology

1983

ABSTRACT

CLARIFICATION OF THE ROLE AND CONFIGURATION
OF LEARNING AS THEY ARE MANIFEST IN
PERFORMANCE AT HALSTEAD'S CATEGORY TEST

By

Kenny William Bertram

Review of the theoretical, clinical, and empirical histories of the Halstead Category Test (HCT) led the present investigator to argue that the test demands diverse and complex activities which recruit virtually all areas of the cerebral cortex. It was also reasoned that the HCT is best viewed as a test of learning. The predictions involved in the present study followed from this hypothesis.

Subtests III through VI were treated analytically as independent records of item response behaviors, and within subtests, items were organized into groups, or sequences, on the basis of structural characteristics. Items were further grouped according to ordinal position within sequences, and by correct option. Item incorrectness was construed as a dependent variable. The aforementioned factors, plus linear, quadratic, and cubic components for Trial and Sequence, their various interactions with one another, and with brain damage, defined by an Augmented Impairment Index, were entered as within groups independent variables. It was predicted that when correct option had been partialed away,

item correctness would increase across trials and sequences.

The neuropsychological protocols of 159 referrals of diverse etiology at a Midwestern Veteran's Administration Medical Center formed the sample for the study.

A multiple linear regression strategy was applied, and the between subjects effect for Augmented Impairment Index proved significant, as anticipated. Within subjects effects for Option, Sequence, and Trial also were significant, and though the results were more complex than anticipated, it was concluded that learning had been demonstrated. The occasional significance of the quadratic and cubic aspects of Trial and Sequence was traced to item characteristics and the incompletely balanced distribution of Option levels across levels of Trial and Sequence.

Interactions were also sporadically significant, and these results were largely attributable to incomplete balance in the design, plus structural peculiarities among the item stimuli.

It was concluded that while learning had obtained as predicted, it was a determinant of item behavior of modest importance. It was speculated that a more balanced analog of the HCT would permit a more definitive evaluation of the study's hypotheses.

ACKNOWLEDGEMENTS

I wish to express my great appreciation, my respect, and my liking to the sage and sane chairperson of my dissertation committee, Norman Abeles. Thanks, Norm, for much useful, helpful, and supportive counsel.

I wish to emphasize my gratitude to Neal Schmitt, a member of my dissertation committee, for having been the major of desperately few real sources of inspiration to me in recent years. When I wasn't busy feeling dull in your presence, Neal, I was learning more than I would have believed possible.

I remain grateful, as well, for the involvement of the other two committee members, John Hurley and Joseph Papsidero, who have affected me lastingly and positively. In particular, thanks for helping me learn to write, John. And Joe, thanks for your longstanding interest in helping me develop myself as a scientist and academician.

With great affection I also wish to acknowledge the guidance and friendship of J. Edwin Mason, my first mentor in Neuropsychology, and probably the person most responsible for my lasting interest in the field. Thanks Ed, and love.

Finally, I seek here to recognize the great influence which Ward Campbell Halstead has had upon my thinking in Neuropsychology, and upon the direction my career has taken,

at this time. It seems to me that this brilliant and energetic man has received but a fraction of the acknowledgement he merits. Being a scientist of the first water, he devoted his years to goals other than material success and acclaim from his colleagues. As such, he avoided letting public attention affect the focus of his energies. It is a pity the same can not be so convincingly said of his theoretical progeny.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF TABLES

LIST OF TABLES (CONTINUED)

LIST OF FIGURES

LIST OF FIGURES (Continued)

Introduction

The scientific objective of this project was to investigate and document the operation of learning, as it is manifested in respondents' behavior at the Halstead Category Test. This purpose, however, was largely incidental.

The stimulus to undertaking this study was, that is to say, not the speculation that the Category Test was a measure of learning, but rather, the investigator's nagging awareness of his own ignorance of precisely, or even vaguely, what the Category Test was a measure of. Worse yet, even those mentors, well-established and practicing neuropsychologists, consulted, could shed but little light on the matter. And so it eventually became imperative to consult the literature in hopes of developing an understanding of the test.

The introductory part of this project, consequently, involves a critical summary of much, though not all, of the research which has, for one reason or another, incorporated the Halstead Category Test. The literature itself quite obviously converged upon the theme of learning in Category Test-taking behavior.

The project can also be viewed as a clear step away from further evaluation of the discriminant validity of the test, a stride in the direction of task-analyzing the test.

Or, it may as well be concluded, once and for all, that the Category Test is a superior indicator of brain damage, and it is high time to begin the process of discovering those perceptual and cognitive elements providing the basis for its capacity to so powerfully discriminate. Obviously, a single study can have but scratched the surface in this regard, but, in addition to a few other, similar efforts which have been made, perhaps an acceptable beginning now exists.

Finally, the project can be thought of as something of a newcomer's eulogy to Ward Campbell Halstead, the brilliant and meticulous scientist who almost singlehandedly developed the currently immensely popular battery which bears his name. It would seem, somewhat in contradiction to the uses to which his tests during the past 35 years have been put, Halstead cared less for his Battery than for that construct they were designed to elucidate, 'Biological Intelligence.' Good for you, Dr. Halstead, and may the current endeavor, in its own humble way, serve as a redirection of attention back upon Biological Intelligence, and as a tribute to you.

History and Development of the Halstead Category Test

In the latter half of the 1930's, Ward Campbell Halstead (1908-1969), having just completed his doctoral work in experimental psychology, had been rigorously pursuing clinical and empirical examination of brain-damaged humans. Though obviously a thinker of particular breadth, he was especially intrigued by what he called "grouping behavior" (Halstead, 1940), or more descriptively, the sorting of objects or stimuli into categories based on similarities and differences in one or more of their characteristics.

Halstead's work in this context was heavily influenced from two directions. The first of these was the thinking and research of the obscure scientist, Heinrich Kluver, originally a mentor and later a colleague to Halstead at the University of Chicago. Kluver's (1929, 1931, 1936) major impact upon Halstead's thinking concerned his "method of equivalent and non-equivalent stimuli," essentially an approach to the study of learning from the point of view of stimulus characteristics. Clearly a learning theorist, Kluver's interest was in determining which aspects of a complex stimulus array were relevant insofar as eliciting a given response from an organism was concerned.[1] He considered the stimuli "functionally equivalent" (Kluver, 1936) when irrespective of their apparent differences, they elicited identical behaviors. Much of his research consisted of the extremely careful study of stimulus differences and similarities in search of precisely those aspects, in a given situation, which accounted for the facilitation or suppression of a conditioned response.

Though most of his work involved animals, and especially higher primates, Kluver (e.g., 1936) did from time to time discuss its applicability to the study of human learning, and even to such examples of abnormal human functioning as psychopathology and brain damage (i.e., to "breakdowns", as it were, in the usual operation of stimulus equivalence or non-equivalence).[2] It is likely that Halstead found these speculations intriguing.

Following Kluver's (1929, 1931, 1936) lead, it is possible to view the operation, within the organism, of stimulus equivalence or non-equivalence, as the activity of abstract thought. Each function, that is to say, involves an ongoing process of solving a problem. Events or objects are in some cardinal sense or senses functionally equivalent or isomorphic, and in many other senses, functionally dissimilar, or distinct. The <u>problem</u> to be solved is that of detecting and cataloging similarities and differences, and of somehow sorting through the comparisons of attributes and selecting the <u>salient</u> similarity or similarities or differences. The solution process is <u>ongoing</u> because a new achievement of functional equivalence or non-equivalence must be sought at the addition of each bit of information.

Inferring a conceptual isomorphism between the notions of stimulus equivalence and abstract reasoning, as Halstead (1939) so obviously did, leads naturally to a consideration of the other important influence on Halstead's formulations: the German neurologist, Kurt Goldstein.

Like Kluver, and very many others of the era, Goldstein's (1936, 1939, 1940, 1942, 1944) thinking was characteristic of the Mentalistic and Structuralistic traditions in that he was more than willing to speculate at length concerning the internal, intrapsychic machinations corresponding with empirically available, behavioral events. During the hiatus between the first and second World Wars, Goldstein and his colleagues intensely studied the victims of brain injury sustained during the first of these Wars, ostensibly (Goldstein, 1940) with the hope of deriving an account of human behavior. The idea (and by no means a new one) was to develop an understanding of normal behavior by investigating the character of its disruption secondary to cerebral insult. Consistent with the Mentalistic and Structuralistic orientations, both the (inferred) internal experience, and the neurobehavioral characteristics, of overt behavior were stressed. Goldstein's work often turned in the direction of the highly inferential and the anecdotal, and he has been somewhat harshly criticized for this (e.g. Battersby, 1956; Reitan, 1958) in more recent times. Nonetheless, a great many of his more prominent theoretical tenets remain explicitly or implicitly popular today (Walsh, 1978, pp. 120 ff.).

Goldstein was particularly fascinated by what he saw as the impact of cortical damage upon the "abstract attitude" (Goldstein, 1940, 1942). This, he defined, as follows:

> The abstract attitude is the basis for the following *conscious* and *volitional* modes of behavior:
>
> 1. To detach our ego from the outerworld or from inner experiences.
> 2. To assume a mental set.
> 3. To account for acts to oneself; to verbalize the account.
> 4. To shift reflectively from one aspect of the situation to another.
> 5. To hold in mind simultaneously various aspects.
> 6. To grasp the essential of a given whole; to break up a given whole into parts, to isolate and to synthesize them.
> 7. To abstract common properties reflectively; to form hierarchic concepts.
> 8. To plan ahead, ideationally; to assume an attitude towards the "mere possible" and to think or perform symbolically.
>
> (Goldstein and Scheerer, 1941, p. 4)

Goldstein envisioned human thought as being capable of but two orientations or modalities: the abstract one, and the concrete one (Goldstein, 1940, 1942). The concrete attitude he defined as being the opposite of the abstract attitude, and hence being characterized by stimulus boundedness, rigidity, an absence of detached consideration or formulation, and so forth. Goldstein also argued that the two attitudes were functionally mutually exclusive, and, moreover, that the function of the abstract attitude was to oversee, and to deliberately and thoughtfully plan and control behavior, which Goldstein viewed as consisting of fundamentally concrete operational units (Goldstein and Scheerer, 1941). As might be expected, Goldstein's position was that injury to brain compromised or even completely abolished, the capacity to adopt the abstract attitude.[3]

Goldstein and his co-workers developed or improved upon several tests of the capacity to adopt the abstract attitude (Goldstein and Scheerer, 1941). The tests share the demand upon the respondent that he or she formulate an explicitly abstract, hierarchical solution to a given problem. For example, one test required subjects to sort skeins of yarn of various shades into groups. Correct test behavior depended upon the formulation of the abstract (i.e. transcendent, relatively intangible) concept, hue. More concrete, and incorrect solutions included sorting according to redness, blueness, etc. (i.e., categorizing on a basis other than hierarchical). Goldstein was in no sense a psychometrician, and he consequently derived only qualitative scoring approaches for his tests. He has been somewhat heavily criticized for this (e.g., Battersby, 1956).

One of the tests Goldstein endorsed, The Gelb-Goldstein-Weigl-Scheerer Object Sorting Test, has been described in some detail by Weigl (1941). Materials consisted of 30 common household objects (e.g., a pipe, candles, bell) selected such that they could be organized or sorted into groups on the basis of one or several hierarchical organizing principles. The objects were placed upon a tablecloth in "standard position," and subjects were directed to sort them into categories of their own design. Again, no quantitative mode of scoring existed, but brain-damaged people demonstrated less facility at the task than did normal controls. In particular, the cortically impaired: (1) tended to place objects together which would be used together (e.g., tools); (2) were often unable to decide that objects might be sorted according to more than a single organizing principle; (3) were typically uninsightful concerning the bases for sorting they were able to arrive at (Weigl, 1941).

The Gelb-Goldstein-Weigl-Scheerer Object Sorting Test is relevant not just because it typifies the means of assessing deficits in abstract thinking which were popular at the time. It was an important ancestor to one of the more widely used measures of brain damage today, the Halstead Category Test, or HCT (Halstead, 1940).

Halstead's (1939, 1940) earlier work with "sorting tasks" employed a test very much like the Gelb-etc., but with 62 objects rather than 30. For an illustration of the standard administration, consult the 1940 Halstead paper. The administration procedure was more psychometrically rigorous and a careful, quantitative scoring procedure was developed. Halstead (1940) was able to demonstrate, with 26 meticulously documented neurosurgical patients and 11 normal controls, that brain-damaged individuals tended to sort fewer objects than intact individuals, and that the presence or absence of frontal lesions "accounted" for the majority of the variance observed (i.e., frontally lesioned patients performed more poorly than those lesioned more posteriorly).[4] In a test of imminent recall, it also was discovered that subjects with brain damage remembered fewer objects than normal subjects, and again, that frontally injured patients performed more poorly than other patients.

Halstead performed a rather thorough qualitative analysis of the strategies utilized in grouping objects by the various intact and lesioned subjects, in hopes of elucidating the operation of Kluver's (1931) principles of equivalence and non-equivalence in normal and brain-damaged functioning. In general, it obtained that frontal patients tended to group fewer objects, and to approach the test with less flexible and less stable organizing principles than did normal or posteriorly lesioned individuals. It might be concluded that judgments of equivalence and non-equivalence were more difficult to formulate for the frontally lesioned subjects, and were approached somewhat uninsightfully. It is important to emphasize, however, the differences here were subtle, and the strategies adopted by lesioned and intact individuals were not really qualitatively distinct, as Goldstein (1939, 1944) might have predicted.

The work culminating in the previously summarized (Halstead, 1940) study apparently stimulated Halstead to develop a more elegant, standard, and controlled procedure (Halstead and Settlage, 1943). This procedure manifested the desideratum of systematically varying certain aspects of stimuli, in this case, "geometric figures," while assuring that other aspects remained immutable. The procedure was the direct ancestor of what currently is known as the Halstead Category Test (Reitan and Davison, 1974, pp. 366-368), and was identical to the contemporary version save that it utilized nine subtests of 40 trials each, for a total of 360 trials, as opposed to seven subtests of various lengths (i.e. from eight to 40 items or trials), and summing to a total length of 208 trials.[5] The reader is directed to Appendix 1, which contains thorough verbal and depictive descriptions of the Halstead Category Test.

A remarkable, largely unprecedented, and, as yet rarely replicated, feature of this completely standard procedure were the demands placed upon the respondent that she or he: (1) develop, as hypotheses, organizing principles, or abstract concepts, with regard to items or trials within a subtest; (2) evaluate these hypotheses in accordance with positive or negative reinforcement; and (3) adjust the hypotheses in an ongoing or dynamic way on the basis of the reinforcements received.[6] It can be readily seen that good performance at the test demands relatively high level skills in noting similarities and differences among stimuli, the capacity to incorporate feedback into ongoing cognitive and behavioral planning and organization, and to some degree at least, a reasonable short term memory (Walsh, 1978, p. 121).

With a group of six carefully described neurosurgical patients, and a group of ten normal controls, Halstead and Settlage (1943) demonstrated that patients with frontal or prefrontal cortical ablations performed dramatically poorly at the test relative to intact subjects. Patients with cortex removed elsewhere were noted to perform about as well as normal controls (but see note 4). The authors, characteristically, inferred that frontally damaged individuals had lost their capacity to evaluate equivalence and non-equivalence of stimuli.

During the middle and late 1940's Halstead and a number of his colleagues were awarded a substantial, federally backed grant for the purpose of explicating the relevance of the frontal lobes to organized complex human behavior. This ambitious endeavor culminated in the completion of Halstead's (1947) book, the only one he ever wrote, and in which appears the most comprehensive account of his theoretical conceptualizations.

The project involved the administration of 27 "quantitative indicators" of neuropsychological and psychological functioning to 207 subjects with known (and, as usual, quite carefully documented) brain damage, and 30 cortically intact controls. From this full array of 27 indicators, 13 were selected for subsequent intercorrelation and factor analysis, because of their amenability to parametric statistical treatment.[7] This set of 13 variables was factor analyzed independently by Holzinger (Halstead, 1945) and by Thurston (Halstead, 1947). Holzinger derived two alternate structural decompositions, one orthogonal, and one oblique. Thurstone apparently applied his principal components extraction (Thurstone, 1947) with a carefully executed oblique rotation. All three approaches yielded four factors. Though he (Halstead, 1945) published Holzinger's solutions, Halstead evidently preferred Thurstone's and this is briefly discussed below.

Thurstone's first factor Halstead (1947, pp. 43-55) labeled "C", or "Central Integrative Field." He believed this construct to involve the matrix of overlearned behaviors and cognitions we all possess, garnered from having distilled the pertinent elements of thousands of experiences and situations. Halstead believed the operation of "C" to be characterized by the elicitation of the stored, relevant material, in response to the new situation, whose critical features are then subsequently integrated, themselves, into this central field. He felt that this "organized experience of the "individual" was roughly coextensive with the psychoanalytic term, "ego" (Halstead, 1946, 1948). He also saw the importance of memory to the Central Integrative Field (Halstead, 1951).

Thurstone's second factor, the "A" or "Abstraction" construct, Halstead believed to underly or drive what he called "grouping behavior" (e.g., Halstead, 1940). Halstead had examined perhaps a dozen different grouping tests and techniques (Halstead, 1947), and carefully documented four distinct forms of grouping behavior (or, four grouping strategies). Without going into unneccessary detail, he felt that these strategies existed at different points along a continuum from what might be called "unaware," or "irrational" abstraction, to what might be thought of as "rational" or "conscious" abstraction.[8] Both involve the selection, from among many, of a single property or aspect common to all members of a class of objects which differ significantly in other regards. Halstead (1951) argued that the operation of "A" obtained when the biologically "wired," or "irrational" abstraction was held in abeyance, and the consciousness sought other hypotheses, or "organizing principles" (Halstead, 1947) by means of which to group or categorize objects or events. He also believed the "A" factor operated to permit the discarding of an organizing principal, and its replacement with a more appropriate or powerful one as the need for this arose.

Halstead (1947, pp. 68-83) named Thurstone's third factor "P," or "Cerebral Power." He believed this construct represented the capacity to willfully direct concentration, to control otherwise disruptive affects or impulses, to delay gratification, and the like. He believed the operation of this factor, unlike the others, to occur as a function of cerebral metabolism.

The final factor Halstead (1947, pp. 84-90) labeled "D" or the "Directional" factor. Its operation might be thought of as the avenue or modality by means of which any of the three other, more process-oriented factors emerges, occurs, or is "exteriorized" (Halstead, 1951). This construct might more clearly be understood as the behavioral or cognitive flexibility with which the other three factors are expressed or put into operation. Halstead (1947) viewed "D" as being particularly salient in situations demanding that the individual adopt unusual modalities for sensory or other activities (e.g., navigating in a completely darkened room).

Together, Halstead believed the processes represented by these four factors produced "Biological Intelligence," the basic function of the central nervous system, and particularly of the frontal lobes (Halstead, 1948). Halstead

(1951) held that the operation of Biological Intelligence was responsible for any and every adaptive and intelligent central nervous system activity, and that Biological Intelligence was the attribute of the individual which was compromised when the cerebral cortex, especially the frontal cortex, was damaged (Halstead, Carmichael, and Bucy, 1946).

Halstead's (1947) Category Test (HCT) loaded solidly upon factors "C" (.49) and "A" (.63). These two factors, though members of an obliquely rotated set, were essentially orthogonal (r = - .02). These results imply that the HCT demands both the careful, volitional abstraction of salient features of (visual) stimulus objects, and the capacity to integrate new, with previously existing, information. This is consistent with what various workers (Reitan and Davison, 1974, pp. 366-368; Walsh, 1978, pp. 294-295) have observed since Halstead's time.

Post - Halsteadian Development of the Category Test

Beyond the late 1940's and early 1950's, Halstead's interests apparently veered away from the business of validating and refining his battery of neuropsychological indicators, and he became consumed instead with the study of the effects of specific conditions (e.g., ablative surgery, hypertension, noise) upon brain function, or Biological Intelligence (Halstead, Apter, and Heimburger, 1951; Halstead and Chapman, 1954; Halstead, Chapman and Symmes, 1955). He remained convinced that the frontal lobes were especially endowed with this quality (e.g., Halstead and

Shure, 1958), even when others began to challenge both his theoretical propositions and his empirical procedures (Chapman and Wolfe, 1959; also see Walsh, 1978, pp. 113-117).

There have been several lines of research pursued with the HCT since Halstead's early work. These include: (1) its further validation; (2) correlating it with other variables of neuropsychological relevance (principally, intelligence and age); (3) attempting to shorten or simplify the test; and (4) developing an account of the measure's psychometric and neuropsychological characteristics. In the main, this conceptual and empirical work has been undertaken, or at very least heavily influenced, by one of Halstead's earliest graduate students, Ralph Reitan, the individual primarily responsible for the refinement and popularity the Halstead battery of tests has enjoyed during the past 30 years. In fact, so much involvement by Reitan has occurred that the battery now is generally known as the Halstead-Reitan. To the extent made pertinent by the objectives of the current project, each of the research focuses listed is briefly characterized, below.

Validity of the Halstead Category Test

The measure's validity has generally meant it's capacity to distinguish between groups of people with and without brain damage, with as few false negatives and positives as can be managed. The model for validity studies of the HCT, and, indeed, the entire Halstead-Reitan Battery, was established by Reitan's (1955a) now classic endeavor, in which he compared 50 pairs of subjects, carefully matched for age, sex, education, and ethnic origin, where one member of each pair manifested, as it were, "proved" (Reitan, 1955a, p. 29) brain damage, and the other remained apparently normal. The Halstead Battery was administered to all members of both groups. Of the total array of ten measures, the Halstead Category Test proved the most accurate discriminator between paired subjects, next to an index, the Impairment Index (Halstead, 1947), based upon all ten measures. With the HCT, only three subjects with cortical impairment produced better scores than their intact counterparts. In a subsequent report Reitan (1956) documented that HCT and the Impairment Index correlate substantially (r = .71 for 50 brain damaged subjects; r = .50 for 50 intact subjects). This not only underscores the validity or _efficiency_[10] of the HCT in this context, it also serves as evidence concerning the complexity of the measure and the demands it exerts upon the respondent's perceptual and cognitive faculties. Indeed, and in contradistinction to Halstead's (1947) frontal lobe "manifesto," it was demonstrated by Reitan (1955a) that HCT performance is rather unequivocally compromised by lesions _anywhere_ upon the cerebral cortex, and, moreover, that the deleterious influence of cortical impairment in this respect increases directly as a function of lesion size (Chapman and Wolff, 1959). The determinants, or more relevantly, the perceptual and cognitive components of HCT performance are clearly complicated, and draw upon many different cortical sites.

A great number of other, nearly identical studies have similarly demonstrated the discriminant validity of the Halstead-Reitan Battery and by necessity of the HCT (e.g., Reitan, 1966, Shaw, 1966; Vega and Parsons, 1967; Russell, Neuringer and Goldstein, 1970; Reitan and Davison, 1974, Filskov and Goldstein, 1974). The frequency of this sort of validity study is apparently increasing (Hevern, 1980), and with the development and wider dissemination of multivariate statistical technology, accuracy in discrimination has improved (e.g., Wheeler, Burke, and Reitan, 1963). The model for all of these studies and a great many more has been that of Reitan's (1955a), after Halstead's (1947) pioneering work. The design may be referred to as an extreme, or static groups approach (Campbell and Stanley, 1963). Of its many faults, one of the most disagreeable is the facilitation of the development and use of measures or entire batteries which are highly valid discriminators, but are entirely conceptually opaque. Diagnosis is rendered definite, but little headway is made concerning the nature of what is being diagnosed. The impairment Index, a value probably useless for any purpose save indicating, with substantial accuracy, the odds of brain damage existing in a given case, best typifies the fruits of this strategy, called, by Rourke (1982), "static neuropsychology". The Halstead Category Test, as both the most important and probably the most complex measure in the Halstead-Reitan Battery, is perhaps the second most grievous offender in this regard. What would seem to be essential is to approach the HCT, and the remainder of Halstead's tests, from the point of reference of psychometric or "task" analysis (Rourke, 1982). Certain inroads have in this regard been made, and it is to these which attention must be directed.

Age, Education and the Halstead Category Test

Reitan (1955b) examined 180 subjects with known brain damage and 101 neurologically intact subjects, and discovered that for cortically normal subjects, age correlated with Impairment Index both for subjects under 45 (r = .54), and for subjects within the age range form 45 to 65 (r = .61). For brain damaged people, the relationships were dramatically weakened (r = .27, 45 to 65). These findings indicate that age and cortical impairment are positively related and that the relationship is amplified by age itself, and attenuated by occurrence of brain damage.

In second study, Reitan (1956a) employed 190 known brain damaged individuals and 116 cortically intact subjects, and demonstrated that both psychometric intelligence assessed by means of the Wechsler-Bellevue Form I) and Impairment Index correlated solidly with age. Correlations between total weighted score (WB-I) and age, and Impairment Index and age were .32 and .37 for the brain damaged group, and .35 and .60 for the non-brain damaged group. As before, brain damage operated to attenuate the relationship between age and Impairment Index.

Reitan (1957) similarly evaluated the relationship between HCT performance and age for a group of 138 "normally functioning, high-level subjects" (i.e., mean education was 16.52 years). A correlation of r = .45 obtained between the two variables. The author inferred that abstraction ability falls off with increasing age. Performance tended to decrease most dramatically after about age 30.

In another study, Reed and Reitan (1963a) demonstrated, with 40 matched pairs of intact and brain damaged subjects (mean age, 28 years; mean education, 11.8 years), and two groups of older subjects (n_1 = 46, mean age of 44.7 years, mean education of 16.7 years; n_2 = 29, mean age of 55.3 years, mean education of 13.9 years) that various neuropsychological indicators were as functional in discriminating between the two older groups as they were in discriminating between the two younger groups. The HCT was one such variable. The authors suggested the possibility that the degenerative impact of age upon CNS integrity was in some way analogous to the acute disruption in CNS functioning caused by brain insult or lesion.

A second study by the same investigators (Reed and Reitan, 1963b) involved groups of 40 young (mean age of 28.05 years, mean education of 11.82 years), and 29 older (mean age of 52.96 years, mean education of 12.45 years) subjects. Both groups were administered some 29 standard measures including the Halstead-Reitan battery) which had been rank-ordered by three judges on a continuum from "heavily dependent on prior experiences" to "most dependent upon immediate adaptive ability (and) complexity of the problem-solving." The HCT was ranked "first" on this continuum (i.e., most dependent upon adaptive ability and capacity to solve complex problems). As well, the HCT proved to more clearly separate the two groups than any of the other measures. It was inferred that: (1) The HCT proved more dependent upon capacity to solve more complex problems than the other indicators (i.e., among Halstead's tests); (2) this capacity is impaired in the older brain, just as in the otherwise damaged brain.

In a similar study, Fitzhugh, Fitzhugh and Reitan (1964) compared groups formed by means of splitting a pool of 283 patients with chronic cerebral dysfunction (i.e. seizure disorders) at the median age (35.5 years) with respect to performance at the array of tests utilized in the previously described study (Reed and Reitan, 1963b), and rank-ordered, as in that study, in accordance with a "problem-solving--experiental background" continuum. Once again, the tests toward the "problem-solving" end (e.g., HCT) of the continuum proved more effective in discriminating between older and younger groups than the "experiential" (e.g., remote memory) tests. These results suggest that although both aging and brain damage impair the capacity to solve novel and complex problems, the two sources of deficits are also at least partially independent. That is to say, cognitive functioning which has been chronically impaired due to brain damage deteriorates, nontheless, in much the same way as normal cognitive functioning, with the accumulation of age. The qualification appropriate to introduce here is that epileptic subjects are an extremely heterogenous lot insofar as absolute degree of brain damage, as measured by the HCT or the Halstead-Reitan Battery, is concerned. Many test within the brain damaged range, yet many do not.

With 50 neurologist-confirmed, brain damaged patients (mean age of 41.7 years; mean education of 10.2 years), and 50 neurologically intact, though hospitalized patients (mean age of 40.8 years; mean education of 11.1 years), Vega and Parsons (1967) correlated various indicators from the Halstead-Reitan Battery with age and education. The HCT, correlating more powerfully with age than any other variable used, produced, for the brain damaged group, coefficients of -.36[11] and .22 between HCT and, respectively, age and education. Analogous values, for the intact group, were -.63[11] and .45.

This effort was cross-validated several years later with samples of 35 brain damaged (mean age 34.6 years; mean education 11.2 years) and 25 neurologically intact but psychiatrically disordered (mean age 33.2 years; mean education 12.2 years) patients (Prigatano and Parsons, 1976). Correlations between HCT and, respectively, age and education, were: -.45 and .29 (brain damaged); and -.42 and .21 (psychiatric). Partialing education had no effect on the relationship between age and HCT performance for either group. Interestingly, the HCT failed to discriminate significantly between brain damaged and psychiatric patients, a finding which has been widely established, and which seems to affect not merely the HCT, but as well, the other tests of the Halstead-Reitan Battery (Watson et al., 1968; Lacks et al., 1970; Lewis, Nelson and Eggerston, 1979).[12]

A final study (Mack and Carlson, 1978) in this vein involved three groups, including 40 young and cortically intact volunteers (mean age 25.03 years; mean education, 15.43 years), 41 aged and cortically intact volunteers (mean age 69.76 years; mean education, 14.05 years), and, 43 presumably neurologically impaired patients (mean age, 41.70 years; mean education, 13.00 years). An analysis of variance revealed that young subjects were superior to the statistically indistinguishable older and brain damaged subjects. When treated as a repeated measures, groups (i.e. age) by subtests (utilizing only subtests III, IV, and V) design, both main effects and the interaction proved significant. The effect for subtests was consistent with the notion that learning occurs as a function of experience with the HCT, as subtest performance improved for each succeeding subtest. The significant interaction was attributable to the fact that young subjects, in contradistinction to older and brain damaged subjects, performed more poorly on subtest V than on subtest IV. This finding was, and remains, unexplained.

In summary, these studies suggest that age, like cortical insult, impairs adaptive functioning and the capacity to address novel and complex problems. Relatively unaffected by age were those skills dependent upon old learning, or "prior experience," and this is in fundamental agreement with Wechsler's (1944) notions concerning skills which "hold" and "don't hold" with the accumulation of age. When the current set of studies is considered in conjunction with the previous discussion of the validity studies, it can also be readily inferred that adaptive ability, complex problem-solving, and the capacity to quickly develop new learning, as assessed by the HCT, require an intact cerebral cortex, and are vulnerable to any event or process which involves the loss of cells. As Halstead (1951) pointed out, Biological Intelligence is characteristic of the healthy nervous system.

The attenuation of correlations noted in the brain damaged groups in several of the studies summarized, but especially those of Vega and Parsons (1967) and Prigatano and Parsons (1976) merits some comment. One possible inference for the phenomenon is that both aging and (any other) brain damage exert a similar influence upon adaptive abilities and complex problem solving, and that the more rapidly developing influence of brain damage "preempts" the more gradual accumulation of influence secondary to normal aging. Statistically, this translates to a reduction in variance and covariance in the brain damaged groups, due to restriction in the range of HCT scores (Magnusson, 1967, pp. 144-147).

Finally, like psychometric estimates of intelligence, the studies herein reveal that education is modestly correlated with HCT performance. This relationship is also somewhat mitigated by the appearance of brain damage.

Intelligence and the Halstead Category Test

In an early study designed to flesh out Halstead's (1947) conceptual entity, "Biological Intelligence," or the adaptive qualities inherent in the healthy central nervous system, Reitan (1956b) correlated the Wechsler-Bellevue Scale (Wechsler, 1944), then still a relatively new measure, with Halstead's Tests, including the HCT. In his customary way, Reitan utilized two samples of 50 subjects each, one with, and one without, proven brain damage, and individually matched in pairs vis-a-vis race, sex, and education.[13] Correlations were computed separately within samples.

The Halstead Category Test (HCT) correlated more dramatically with Verbal IQ (r = -.58, brain damaged; r = -.65 non-brain damaged), performance IQ (r = -.64 brain damaged; r = -.67, non-brain damaged), and Full-Scale IQ (r = -.65 brain damaged; r = -.72, non-brain damaged) than did any other indicator in the battery (including Impairment Index), except Speech Sounds Perception Test, which was roughly on a par with the HCT. In contrast to what was routinely observed in relationships between age and HCT (e.g. Vega and Parsons, 1967; Prigatano and Parsons, 1976), very little attenuation in the HCT-IQ relationships obtained in the brain damaged group, relative to the intact group. This would seem to suggest that while age and brain damage manifest overlap in influence, psychometric intelligence and brain damage (or biological intelligence) do not, insofar as HCT performance is concerned.

With the same groups of subjects, Reitan (1959) subsequently demonstrated that the Halstead Impairment Index more effectively discriminated between brain damaged and non-brain damaged subjects than did any of the Wechsler-Bellevue Subtests or the three IQ estimates. It bears mentioning that all of the Wechsler variables, save the Digit Span subtest, also significantly discriminated between the groups. Also of interest, the presence of brain damage appeared to exert a more deleterious influence upon subjects' scores for Wechsler's (1944) "don't hold" subtests relative to his "hold" subtests. The "holds" of course, "hold" their own with age. "Holds" emphasize old learning; "don't holds" the capacity for new learning.

With 29 more or less neurologically normal subjects, Shore, Shore and Pihl (1971) obtained a correlation of -.87 between the Wechsler Adult Intelligence Scale's (Wechsler, 1955) age-equated sum of scale scores and the HCT. The authors also correlated Cohen's (1957) factors for the WAIS with HCT, and obtained coefficients of .84 with Verbal Comprehension, .72 with Perceptual Organization, 1.00 with Memory, and .76 with General Intellectual Functioning. These results would seem to indicate rather more substantial overlap between tested, or psychometric intelligence, and Halstead's Biological Intelligence, as represented by the

HCT, than was discovered by Reitan (1956), in working with the Wechsler-Bellevue Scale (Wechsler, 1944). Note also the rather astonishing correlation between HCT and Cohen's Memory Factor, a variable defined by the simple linear combination of Arithmetic and Digit Span Subtests.

With two samples (n's of 177 and 62) of epileptics, Lin and Rennick (1974) replicated both Reitan's (1956, 1959) and Shore's, Shore's and Pihl's (1971) endeavors. Thus, HCT total error score was correlated with estimates of Verbal, Performance, and Full Scale Intelligence, all component subtest scores, and three of Cohen's (1957) factors. For larger and smaller samples, respectively, correlations were obtained between HCT scores and Verbal IQ (-.51, -.68), Performance IQ (-.60 -.49), Full Scale IQ (-.59, -.65), Cohen's Verbal Factor (-.46, -.68), Cohen's Perceptual Factor (-.61, -.44), and Cohen's Memory Factor (-.48, -.55). Results, in general, were like those obtained by Reitan (1956), and in contrast to Shore, Shore, and Pihl (1971), relationships between HCT and Cohen's factors were more modest. In view of the size of the Shore, Shore, and Pihl (1971) sample (n = 29), the results of Lin and Rennick are perhaps to be the more trusted.

Landsell and Donnelly (1977) factor analyzed the WAIS subtests plus the HCT and the Halstead-Reitan Finger Oscillation Test, a measure of motor speed often useful in lateralizing brain damage (Russell, Neuringer, and Goldstein (1970). Subjects (n = 94) included depressed and other psychiatric patients, and epileptic and other (unspecified) neurological patients (mean age, 39.5 years; Mean IQ 105.6). Their principal components, varimax-rotated solution resulted in four factors, the last two of which were rather minor. The first and largest factor was determined in general by the WAIS verbal subtests, and was labeled, "verbal comprehension." The second, "visuomotor" factor, was dominated by HCT (loading of .82), and included all of the WAIS performance subtests, save Digit Symbol, which, in combination with the Finger Oscillation Test, comprised the fourth, tiny factor, "manipulative speed."

The authors inferred that the HCT does not involve a skill distinct from nonverbal intelligence (i.e., as measured by the WAIS). Given the rather unusual character of their sample, however, this conclusion may be a bit premature. Reitan (1956) discovered substantial correlations between HCT and the verbal Wechsler-Bellevue subtests, which the current authors did not. As well, Shore, Shore, and Pihl (1971), and lin and Rennick (1974) discovered substantial correlations between HCT performance and Cohen's (1957) Verbal Comprehension Factor, as well as its components, the verbal subtests, Information, Comprehension, Similarities, and Vocabulary, plus the overall estimate of Verbal Intelligence.

In summarizing the studies relating psychometric intelligence to HCT, it is perhaps appropriate to observe that while all the information is not yet available, the HCT draws substantially upon both verbal and nonverbal aspects of intellectual functioning. Phrased alternately, the HCT is a complex and demanding task, and it demands diverse cognitive operations for successful performance. There exist some indications that the performance, nonverbal, or spatial manipulative characteristics of intellectual activity are peculiarly relevant. This is not terribly surprising, in that performance/nonverbal skills are least dependent upon old learning, are most likely to succumb to the influence of aging, and generally require the novel solution of complex problems. However, as has already been discussed, the determinants of HCT performance are quite complex, and draw upon both verbal and nonverbal aspects of cognitive activity.

A final comment demanded here is that the WAIS subtests are, themselves, complex tasks which have not, as yet, been adequately analyzed in the interest of revealing component perceptual, cognitive, and motor elements. This renders difficult attempts to meaningfully relate HCT with them.

Additional Studies Concerning the Halstead Category Test

As has been summarized above, it was Halstead's (1947) impression that his Category Test demanded, from the individual, the ability to volitionally engage in "grouping behavior" based upon a careful consideration and selection (from among potentially several foils) of the appropriate characteristic or characteristics of arrays of visual stimuli. The correct aspect or aspects he referred to as the "organizing principle" of the array. In order to emphasize the volitional, or the detached and analyzing, aspect of grouping behavior, Halstead incorporated reinforcement (buzzer and bell) contingencies. The utilization of the feedback mechanism adds significantly to the complexity of the task and, as well, to the demands placed upon the subject, as it requires the respondent to allay responding in order to consider and integrate new information, the impact of which may involve shifting the response set or organizing principle.

Halstead (1947) was in substantial agreement with Weigl (1941) and Goldstein (1940, 1941, 1942), in that he believed that it was the capacity to engage reflectively in tasks requiring organization of stimuli on the basis of abstracted features that: (1) most characterized frontal lobe activity; (2) most effectively distinguished the cortically intact person from her or his brain damaged counterpart. Goldstein (1941) held that this distinction was qualitative, or rather that brain damage abolished the capacity to volitionally detach the focus of attention from a particular object (either external or internal in locus), and to analytically consider several such objects. It is unclear that Halstead entirely agreed with this qualitative interpretation. It is more likely that he viewed abstract (i.e. grouping) behavior as a continuum, bridging Goldstein's (1941) polar opposites, the abstract and concrete "attitudes." Indeed, at one time Halstead (1940) attempted to document and discriminate between different forms or strategies of object sorting, which he apparently believed reposed at various points along

Goldstein's hypothetical continuum.

Reitan (1958) felt the issue remained confused as to whether cortically intact and brain damaged renditions of behavior were qualitatively discrepant or not. In this early study, he demonstrated that median intercorrelations between measures of the Halstead-Reitan battery did not differ significantly between brain damaged and normal groups. In a second study, Reitan (1959) demonstrated that groups (*n* = 52, each) of brain damaged and intact subjects, matched for race, sex, chronological age, and education, both improved in performance on HCT subtest VI, relative to HCT subtest V, which manifests the same "organizing principle." The inference was drawn that new learning can occur even among those with "proven" brain damage. Reitan, further, argued that although brain damaged individuals may be *quantitatively* less able to abstract than normal controls (this, of course, is the basis for the claim that the HCT can discriminate between groups of brain damaged and intact subjects), there exists no difference in *kind*, or *mode* of reasoning utilized. Reitan's inference is something of a presumptuous simplification, as it is impossible, upon the basis of HCT score alone, to determine whether or not the two groups utilized identical information-processing strategies, and varied solely in the degree of efficiency with which they did so, events which would seem essential to conclude qualitative identity.

Those findings were replicated and extended with a subsequent (Doehring & Reitan, 1962) investigation, in which it was shown that although brain damaged individuals score more poorly on each of the HCT subtests than intact individuals, the distribution of errors across subtests is essentially identical for both populations. Thus, both populations manifest maximal errors during Subtest III, and steadily improve across subsequent Subtests. These findings imply that the rates of relative familiarization, or learning, with the HCT, are similar for brain damaged and intact subjects.

In this same report, the authors demonstrated that although patients with right hemisphere lesions performed more poorly at the HCT than patients with lesions to the left hemisphere, the difference failed to attain significance. This latter, marginal discovery was in disagreement with earlier work by McFie and Piercy (1952a, 1952b) who, using one of Goldstein's (1941) special sorting tasks, determined that impairment in abstraction ability was more often associated with left than right hemisphere damage. With grouping tasks other than the HCT, Halstead (1940) found lesion lateralization to be irrelevant. However, Halstead and Shore (1958) found left hemisphere damage to be slightly, though not significantly, more predictive of HCT impairment than right hemisphere damage. Chapman and Wolff (1959), in their careful and detailed analysis of performance, found the opposite to be true. Reitan (1960) found dysphasic patients, and those without dysphasia, though brain damaged, to perform equally poorly with the HCT. Finally, Doehring and Reitan (1961) discovered left visual field defects to be more predictive than right visual field defects of poor HCT performance. This finding suggests that disruption of the primary visual radiation, occipital lobe, right hemisphere, harms HCT success.

As previously, it is probably reasonable to conclude that HCT performance relies upon perceptual and cognitive processes of sufficient complexity that virtually all aspects of the cerebral cortex are utilized. The studies of this section also document quite nicely that: (1) the HCT does require new learning for successful performance; (2) this is true for both intact and brain damaged subjects. The controversy between Reitan and Goldstein was not, of course, be resolved. There exists no way of determining, with recourse only to HCT responses, whether or not brain damaged and intact respondents arrived at solutions (organizing principles) in the same or in different ways, or, for that matter, if upon consistently applying the correct organizing principle, they do so with equal insight or lack thereof.

Formal, Psychometric and Other Characteristics of the Halstead Category Test

Discussion in this section is concerned with the few extant psychometric appraisals of the HCT, with the various tactics which have been utilized in order to render it easier or quicker to administer, and, finally, to the thorough conceptual analysis Simmel and Counts (1957) have graciously afforded it. The methodological criticisms and summarizing comments appearing herein will be seen to converge upon the substantive point of the current project.

The Halstead-Reitan Battery is well known to be long, expensive, and tedious to administer, score, and interpret (Erickson et al, 1978). This has encouraged a number of clinical researchers to somehow shorten the battery (e.g., Golden, 1976; Erickson et al., 1978; Mezzich and Moses, 1980; Barrett, Wheatley, and Laplant, 1982). Problems have arisen anew, however, in that shorter forms inevitably have led to the discarding of entire tests, producing not just a quantitative reduction in information, but a qualitative loss, as well. Additionally, few if any of the shortened versions or (more fashionably) "screening batteries" have been validated with independent samples.

Another strategy interesting those seeking to render the Halstead Reitan Battery more time and cost efficient has been shortening various subtests (e.g., Golden & Anderson, 1977; Calsyn, O'Leary, and Chaney, 1980; Gregory, Paul, and Morrison, 1979). In this regard, there has been particular emphasis on Halstead's Category Test (HCT). Probably, this has been due to the HCT's lengthy administration (i.e., up to perhaps one hour with the incapacitated) and the substantial level of frustration subjects are frequently required to endure (Luria and Majovski, 1977). The test also has the virtue of being the better single indicator of brain damage among the lot of Halstead's tests (Reitan, 1955), second only to Halstead's Impairment Index, a summary quantity dependent on seven indices, including the HCT. In addition, the test is highly reliable (Shaw, 1966; Matarazzo et al., 1976)[14] and has been shown to correlate solidly with magnitude of cortical lesion, irrespective of location on the cortex (Chapman and Wolff, 1959).

Various approaches to designing a less noxious version of the HCT have been explored. The test has been shortened, both with attention having been paid to the fact that item scores are proactively dependent (i.e., HCT performance is dependent on learning, and it is appropriately viewed as a time process), and without. Thus, Kilpatrick (1970), Boyle (1975), and Gregory et al. (1979) have derived shortened versions, more or less item-analytically, which involve items taken out of sequence. Kilpatrick and Spreen (1973) have applied the same strategy, and with far more consideration to psychometric principles, to a version of the HCT Reitan (1974) has modified for use with children 9 to 15 years old. These authors also standardized their shortened version.

It is difficult to determine the utility of these shortened versions for two reasons. First, they have not (with the exception of the Kilpartick and Spreen effort) been restandardized or independently validated. Second, very little is known about the role learning plays as a determinant of HCT performance, and consequently it is impossible to conclude whether taking items out of sequence will dramatically affect HCT test behavior or the ranking of individuals based on total HCT score. As well, these versions remove more items from certain subtests than others, apparently because it has been determined that HCT subtests are not all equally discriminative of brain damage (Boyle, 1975).[15] Yet the removal of many items from a given subtest may exert an effect upon items in subsequent subtests manifest only upon independent validation. Again, the role and the process of learning, as they develop during administration of the HCT have not yet been studied, and it consequently remains difficult to comment definitively upon these shortened versions.

Calsyn (1980) has derived a short version (108 items) which, so he claims, does not take items out of sequence. The version consists of the first four (of the total, seven) subtests. this form has been independently validated (Golden et al., 1981) with promising results. Again, however, though the shortened form held up during this validation, it remains difficult to compare it with the previously mentioned forms, which have not been so validated. Morever, Calsyn's (1980) approach was not psychometrically based: he merely divested the test of the last 100 of its items (or the last three of its subtests), without concern for differential item and scale validities.

Another approach which has been applied to the problem of the HCT is the strategy of deriving a version unchanged in length or order of trials[16], but altered in administration so as to render the test more palatable to clinician, subject, or both. Thus, Beaumont (1975) developed an online program which administers and scores the HCT.

Essentially, no examiner is required with this procedure, which also provides feedback (i.e., buzzer or bell) with invariable latency and is, of course, errorless. The problem, however, is that in general, and in particular with brain damaged subjects, it is essential sometimes to provide ongoing coaching to those being examined. As was said before, this is because the test is highly frustrating. continual encouragement is often required to guarantee that the subject's best effort is being elicited. Indeed, the vulnerability of the test to motivational and affective influence may be part of the reason why HCT errors are predictive, not just of brain damage, but also of various forms of psychopathology.[17] In any event, there exists no means by which an on-line computer can, with appropriate flexibility and judgment, coach the subject who is having difficulty with the test and who is thus in danger of "giving up."

Another administration approach which has recently been developed involves the use of a latent image transfer sheet which provides subjects with visual information concerning the correctness or incorrectness of their responses (Wood and Strider, 1980). In a design which was counterbalanced for order of administration, visual feedback was alternated with the traditional auditory feedback procedure, from subtest to subtest. No significant differences were noted, although the effect for test form was obviously confounded with deviation from standard administration (i.e., what were being compared were two alternate forms, both different from the standard HCT). Morever, even assuming that their comparisons <u>were</u> methodologically reasonable, the subjects involved were psychiatric, rather than brain damaged patients. As was pointed out above, performance on the HCT is powerfully influenced by the presence of psychological disorders, for reasons as yet very poorly understood. It would seem inappropriate to generalize results based upon this sample to either "normal" or brain damaged populations. Finally, even assuming that these problems are trivial, their samples were small (i.e., two groups of 25), and the power of their tests was low (i.e., on the order of .30, assuming alpha of .05, and a two-tailed test), rendering it likely that even if the test forms <u>were</u> different, the authors would not have been able to detect it in the first place. Subjects in their (Wood and Strider, 1980) study were questioned, and generally preferred standard administration, because it did not require that they look away from the projected images (i.e., test stimuli) in order to respond.

Adams and Trenton (1981) utilized an identical visual feedback procedure, but supplanted the projector in their alternative administration with a deck of 3" x 5" cards with the HCT stimuli printed on them. Their professed aim was the development of a form of the HCT which would permit group (i.e., relatively unsupervised) administration. With two groups of 30 "normal" subjects each, the same, counterbalanced design of Wood and Strider (1980) was adopted.

Split-half reliabilities were computed for these groups (r = .79), and for a third group of 100 subjects representing an unspecified population who took the standard HCT (r = .82), and the reliabilities were found not to significantly differ. No means or standard deviations were reported for any of the three groups. The authors stated an interest in determining if test halves which differed in administration format would correlate differently. One wonders why the authors concerned themselves so exclusively with this line of inquiry, as it so clearly limits the relevance of their findings, whatever _they_ were. At any rate, the same set of criticisms apply here as was the case with the Wood and Strider (1981) study. As well, the comments made earlier concerning the importance of coaching and involvement on the part of the examiner militate against the utility of a group form of the HCT, so long as psychometric equivalence between forms has not been vigorously established.

Kimura (1981) has independently developed a card form of the HCT, in this case with stimuli printed upon 4" x 6" cards. Subjects are asked to verbalize item responses, and the administrator provides verbal feedback (i.e., says "right" or "wrong"). This form was compared with the standard administration with two groups of 15 neuropsychological referrals, and the two forms were found not to differ significantly. Power of the test applied to total HCT scores was effectively zero. Thus, assuming the test forms, _were_ different, it would have literally been impossible to statistically demonstrate this. Group means for total HCT score were on the order of 80, implying that the samples faithfully represented the population of usual neuropsychological referrals. A third group of 15 "neurologically impaired" subjects were administered both forms (standard form first) and compared with a group of 11 "neurologically impaired" subjects with the reverse order of administration. Test-retest correlations were essentially identical for both groups ($\underline{r}$ = .94, slides first; $\underline{r}$ = .96, cards first). These estimates of reliability were also nearly identical with those which have been reported elsewhere under varying circumstances (Matarazzo et al, 1976).[14] The same criticisms mentioned in conjunction with the test-retest approach used by Adams and Trenton (1981) apply here, as well.

Finally, McCampbell and DeFilippis (1979) have developed a "booklet form" of the HCT. This version involves subjects pointing to a number (i.e., one through four) to indicate their response choice for item. Verbal feedback ("correct" vs. "incorrect") is supplied by the examiner. In a preliminary report the booklet and standard form were compared with a counterbalanced design across testings with two groups of 15 college students each. Results of their two-way analysis of variance revealed a significant practice effect (i.e., the replication factor), but a lack of significance in the difference between forms, and the absence of a significant interaction (i.e., order of administration by practice). The test-retest correlations irrespective of order, were high ($\underline{r}$ = .89, slides first; $\underline{r}$= .95, booklet first), as might be expected (Matarazzo et al, 1976).[14] The same basic criticisms apply here. Power of their F-tests, for alpha of .05, was on the order of .15, implying that a significant difference probably would not have been detected, even if one existed. As well, "normal" undergraduate students served as subjects, as opposed to brain damaged individuals.

In summarizing these attempts to reduce the length of the HCT and/or render the test less noxious or expensive to administer, there are several pertinent criticisms. First, the derivation samples have been small, thus guaranteeing an inadequate test of the hypothesis that the two forms of administration are the same. Rather, the investigators are siding with the null hypothesis, since test significance is partly a function of sample size. Naturally, this circumstance has the effect of reducing the degree of trust one has in the results.

Another way of expressing the same problem is that beta (the likelihood of erroneously accepting the null hypothesis) is extremely large, and power (1-beta), the capacity to reject the null when it is false, extremely small. The researchers uniformly ignore this because the "wrong" null hypothesis is being focused upon. In deriving an alternate form for a test, it is more appropriate to test the hypothesis that the two forms are identical against the null that the two forms are distinct.

Modality of administration and feedback are other stimulus aspects which jeopardize the psychological equivalence of forms. It remains unknown as to how significant these changes are, insofar as affecting what the test measures.

Independent validation has been another vexing problem, in that it was appropriately carried out only in the case of the Golden et al (1981) study of Calsyn's (1970) abbreviation of the HCT. This issue has in general been exacerbated by alternate forms having been derived with "normal" (i.e., essentially cortically intact) subjects. Thus, what one possesses is a test of unknown comparability to the original, which has been neither derived, nor validated as an indicator of brain damage. Though Reitan (1958, 1959) has emphasized that brain damage exerts a quantitative, rather than a qualitative, impact upon HCT performance, this does not obviate the necessity to examine the possibility that level of brain damage interacts with test form or test length.

The test-retest strategy of comparing alternate forms is suspect, too. This is because systematic differences in scores have no impact upon correlation coefficients (Cronbach and Gleser, 1953; Cronbach, 1953).[17] As well, the capacity to store and retrieve information has been found to be inversely related to brain damage (Wechsler, 1945; Russel, 1975), and consequently test-retest correlations between alternate forms may themselves vary as a function of brain damage, thus rendering this index of equivalence in forms moot.

The issue essentially reduces to the matter of deciding what aspects of a stimulus array can be altered while maintaining some assurance that the response (both internally and externally, or empirically) is identical. As Brunswik (1956) has observed, this in itself can (and, from the point of view of valuing a scientific account, should) be rigorously studied. As has been emphasized, in a test of this type, the problems and confounds resulting are numerous and egregious. The most expedient means of resolving these problems would be simply to treat the alternate versions of the HCT as qualitatively distinct and potentially useful indices of brain damage and then to appraise them by proceeding in the usual way, namely attempting to predict brain damage with them, and relating them to other indices (including the standard HCT) of brain damage. In this way, the two salient questions, concerning equivalence of forms, and capacity of alternate forms to detect brain damage, can be directly and appropriately addressed.

Consideration of those forms involving the discarding of items presents additional problems, which, though obviously tractable by the sensible methodology just outlined, suggest several interesting questions. First, as the very description (Halstead, 1940, 1943) of the HCT implies, the capacity evaluated by the test is that of learning. The test, that is to say, is not a test of power (Anastasi, 1976), in the usual cross-sectional sense. Rather, the test appraises the individual's capacities in a dynamic, or longitudinal way.[6] Conceptually, this is a minor problem in that it is not difficult to imagine a single, total score which represents the dynamic capacity of an individual to solve an ordered set of rather complex, yet formally similar problems. However, since the capacity is a dynamic one, the ordering of items is of critical importance. Without inviolate ordering, and unless it can be assumed that items are functionally equivalent with regard to one another, and that the length of the test does not interact with the learning process, then forms of the HCT based upon the discarding of items are of dubious equivalence with respect to the standard HCT.

The first of these assumptions, that the items are functionally equivalent, does not obtain in the HCT (Simmel and Counts, 1957). Rather, it seems that some items are more difficult than others, partially because of their stimulus characteristics, but also because of their position in the ordering within a subtest, *and* because of the interaction between the two factors, stimulus characteristics and position in the ordering. Thus, not only are item responses not experimentally independent, but their stochastic dependence is influenced by their formal characteristics. In actuality, this is saying no more than that item difficulty, in the usual sense, varies in the HCT, but since HCT performance is a time process of sorts, item difficulty is not just a "main effect" insofar as test performance is concerned, but also it creeps into an interaction effect.

The second assumption, that length of the test does not interact with the learning process, is clearly invalid. It is well known that at least up to a certain point, learning, as defined by the change in the probability of a response or a correct response, as a function of time, is not a linear process (Rachlin, 1976, pp. 180-190). Shortened versions of the HCT will likely tend not to be only linearly related to longer versions, assuming that the test measures learning.

Simmel and Counts (1957) have completed the most meticulous and thorough evaluation of the HCT to date. Their sample included 35 neurological patients, all but three of whom manifested psychomotor seizures with involvement of the anterior temporal lobe, either hemisphere. As well, 26 student nurses were included as their control group.

Rank-order correlations computed between item number, within subtests, and number of subjects obtaining a correct response, suggested that performance improves as a function of familiarity with the subtest. Thus, for subtests III, IV, V, and VI[20], correlations, respectively, were .35, -.03, .35, and .10 for the sample of patients, and .70, .55, .41, and .08 for the normal group. As has consistently occurred, correlations were slighter for the impaired group. The authors also noted that although the likelihood of correct responses increased within subtests as a function of item number, the graph of the relationships hardly approximated the usual learning curve, because items are arranged in sequences or clusters of highly similar characteristics, within subtests. Between last and first trials, or items, of adjacent clusters, the authors (Simmel and Counts, 1957) observed precipitous drops in probability of correctness. The emerging curves were upward treading, but sawtoothed, because of this intercluster phenomenon. Perhaps because of the irregularity of the process the authors made no attempt to further study or evaluate the operation of learning in HCT behavior. These investigators also obtained test-retest correlations over a three month interim of .70 for 21 temporal lobe patients, .74 for 26 student nurses, and .80 for all 47 individuals. For 20 patients, test-retest reliability was estimated at .88 over a hiatus of 15 months. These values are not unlike those reported by Matarazzo et al. (1976).[14] The authors also noted the tendency for all subjects to manifest improved performance over time, again suggesting that learning was occurring.

Kuder-Richardson (or, Cronbach's alpha, if preferred) estimates of internal consistency[21] were also computed, and revealed values of .96 (subtest III), .96 (subtest IV), .89 (subtest V), .91 (subtest VI), .75 (subtest VII), and .96 (subtests III through VII, and I through VII). These values closely approximate the split-half estimate of reliability found by Shaw (1966).[14]

The remainder of this (Simmel and Counts, 1957) monograph was devoted to a careful scrutiny of the distributions of respondent's errors and their relationships to item characteristics. Their work is both esthetically and scientifically pleasing, though not of great relevance to the current discussion. Certain details of their work will be referred to, as appropriate, in the presentation of the methodology. Intact and lesioned subjects were found not to differ appreciably insofar as the patterning of correct and incorrect responses was concerned, although the impaired sample fared more poorly at the test.

Summary and Outline of the Focus of the Current Endeavor

What has been covered thus far tends to converge upon the thesis that the HCT assesses a complex of phenomena having to do with abstract and conceptual thought, and the flexible solution of novel and complicated problems. HCT performance is exceedingly vulnerable to any form of cortical insult, demands substantial verbal, and perhaps especially nonverbal, intelligence, and is quite dramatically impaired by the accumulation of age beyond about 30 years. It is tenable to argue, consequently, that the HCT is acutely sensitive to the loss of cortical tissue occurring for any reason, and probably this is because the HCT demands the use of most aspects of the cortex, or, back to the original statement, the test assesses a complex of phenomena. It is here argued, perhaps more out of vehemence than originality, that the complex process evaluated by the HCT is nothing more mysterious than the operation of learning, the acquisition of new ways of organizing information, the formulation of new observational sets or proclivities, the forging of new connections between stimulus and response. Successful HCT performance is associated, then, with the ability to *intentionally* apply abstract thought *and* the capacity to integrate new information with old (Halstead's, 1947, "Central Integrative Field"), and also to incorporate responsestrengthening and response-weakening information into one's cognitions, and ultimately, one's behavior.

It also is believed possible to conduct an exploration of the appropriateness and the veridicality of these assertions by means of studying, in the tradition established by Simmel and Counts (1957), the internal characteristics of the HCT.

Actually, two connotations of this notion, "internal characteristics" may be appealed too. The first sense is the traditional conception of internal structure (Cronbach, 1951), which emphasizes the factorial composition of tests or subtests on the basis of the item covariance (or correlation) structures. The aim is to derive a coherent image of the sources of variance important in producing the distribution of subtest or total test scores. The second sense is the characteristics of the item stimuli themselves. In general case, the two versions of internal characteristics are more or less identical, because items can be reasonably fully represented by their difficulties (i.e., probabilities of correct responses), and their correlations, with one another, with total test (or subtest, as appropriate), and/or with some remote criterion. In the case of the HCT, it is argued that these aspects of test behavior are important, but it is also true that HCT items are unusually amenable to analysis and summary on the basis of their formal characteristics. This is because they are visual arrays which have been carefully and systematically designed and juxtaposed.

Prior to considering the hypotheses and focal predictions in any detail, it is essential that the reader become familiar with the structural characteristics Halstead (1943, 1945, 1947) has designed into the HCT. In no general description of this test (e.g., Reitan, 1955; Reitan 1966; Reitan and Davison, 1974, pp. 366-368) are these characteristics emphasized or made apparent. Yet they are of critical importance in any attempt to evaluate the internal characteristics of the test.

The reader is referred to the Appendix, in which the various stimuli (items) for subtests III through VII of the HCT are illustrated. The tables contained in this appendix have been adapted from Simmel and Counts (1947). Subtests I and II are not considered here, because, as Simmel and Counts (1957) have pointed out, item variances for these subtests are diminishingly small, since virtually everyone achieves a perfect score.

From Figure I-C, it can be seen that the first 32 items are organized into eight four-item clusters. Within clusters, the stimulus arrays are identical, save that the correct response migrates from item to item. The last eight items are not organized into clusters, but rather the stimulus array shifts after each item. The "organizing principle" for this subtest is that the most dissimilar of four geometric figures is the correct response.

Items may also be organized by the option which is correct. Each of the four options is used ten times in the 40 items, and sequencing of correct options appears to have been randomly selected. This is a critically salient attribute, in that throughout the test, some options are far more likely to be chosen than others (Simmel and Counts, 1957). This is because subjects tend to rely upon <u>counting</u> in preference to more complicated strategies for deriving item responses. This, in turn, is partly because subtest II embodies counting of figures as the <u>appropriate</u> "organizing principle", and once having established this set, subjects are loathe to abandon it (Simmel and Counts, 1957). As an example of the ways in which counting operates in effecting item responses of one variety or another, consider item number five, subtest III. There are four objects (suggesting "four" as a response), arranged as three of one shape (suggesting "three" as a response) and one of a different shape (suggesting "one" as a response). None of these responses is correct, but all can be derived on the basis of a counting rationale. The rationale for considering correct option as a determinant of test behavior is that various counting biases account for an appreciable number of the correct (and incorrect) responses made, across items.

Before continuing with a presentation of the remaining subtests, it is critical to point out that subtest III is special. That is, it is generally the first subtest which gives subjects much difficulty.[22] Consequently, it may be viewed as the first opportunity for learning to occur. This subtest, in a very real sense, displays the subject's baseline test behavior and early departures from it.

The "organizing principle" for subtest IV is the quadrant, in achromatic figures, which is either deviant or missing. This is in part analogous to the preceding subtest, in the sense that deviance (or difference) is the key. However, in this case, the correct response is based on a quadrant schema, with quadrants numbered from one to four, in a clockwise direction, beginning with the upper left quadrant. (Subtest III was based upon ordinal position of the distinctive member of four geometric figures). Subtest IV is visually summarized in Figure I-D, in the Appendix.

Stimuli in Subtest IV can be organized in various ways, in a fashion analogous to that used with subtest III. First, items are arranged into ten clusters of varying length (i.e., from three to six items). As in subtest III, the four response options were randomly assigned to items, and balanced, by guaranteeing that each option is the correct one on ten occasions.

Subtest V is organized on the basis of the proportion (i.e., one, two, three, or four fourths) of a figure which is composed of solid lines, as opposed to broken, or dotted lines. Figures are achromatic, and some are solid geometric figures, while others are merely lines or line segments. This subtest is depicted in Figure I-E, in the Appendix.

Items are organized into six clusters, of length varying from three to nine items. Correct option is distributed as in the preceding two subtests.

Subtest IV utilizes the same organizing principle as subtest V, and may, in a sense, be viewed as a continuation of that subtest. Stimuli are less regular in this subtest, particularly toward its end. Subtest VI is illustrated in Figure I-F, in the Appendix.

The first six items are either identical to items in the preceding subtest, or unused representatives of cluster sets appearing in the preceding subtest. Items 7 through 30 are arranged in three eight-item clusters. Items 31 through 40 are quite explicitly unrelated to the other items in the subtest, though, of course, the same organizing principle is utilized. Correct response option is distributed as before.

Subtest VII consists of items drawn from subtest II through VI. In some cases, items are identical replications of previously seen items. In other cases, items may be viewed as previously unused representatives or members of cluster sets which were employed in other subtests. Subtest I is not represented in subtest VII. Each of the other subtests, II through VI, are represented equally (i.e., four items per subtest), except that subtest V is more heavily represented than subtest VI, with a combined total of eight items. Halstead (1943; 1945; 1947, p. 59) describes this last subtest as a test of "recognition" (i.e., items have been at least *generically* seen before). As far as the current investigator knows, it has not been correlated with measures of memory, though very clearly, at least one of the functions evaluated by this subtest is memory.[23] This 20 item subtest is summarized in the Appendix, Figure I-G. There is no particular organization applicable to subtest VII, save that correct option number is distributed roughly as in previous subtests, except that option two is represented but four times, and option one, six times.

Hypotheses

As argued above, the current investigator proposes that learning is a relevant determinant of successful performance at the HCT. For analytic purposes, and as is spelled out below in a detailed way, "successful HCT performance" was operationally defined as making a correct response to HCT items. Given that learning must imply the acquisition over time and under systematic and relatively coherent environmental reinforcement contingencies, of behavior converging upon nearly continuous success, it was felt reasonable to operationally construe "learning" as the _increase_, across succeeding items, and _within_ a given HCT subtest, in the probability of a correct item response. This, then, was the hypothesis of the current endeavor: Learning, as assessed by the positive change, over time, of the likelihood of responding correctly to HCT items of similar content, is expected to characterize arrays of items _within_ HCT subtests, or alternately, subjects' behavior in response to succeeding items _within_ HCT subtests. The specific predictions selected for evaluation of this hypothesis are outlined in detail in the section concerned with design and analysis. A second principal hypothesis, and one which clearly follows from the original derivation of the HCT was that brain damage, or cortical impairment, as defined by or inferred from performance at the Halstead-Reitan

Neuropsychological test Battery for adults would exert a negative impact upon HCT success. It was also hypothesized that brain damage would impair the capacity to learn.

Method

Subjects

Subjects were 159 individuals referred for neuropsychological evaluation to the Psychology Service of a nearby Veterans Administration Medical Center. All were referred as inpatients. They proved rather a heterogeneous group in most regards, including intelligence, age, educational background, and severity of cognitive impairment. A substantial minority had previous and/or coexisting diagnoses of functional illness, with a history positive for psychological involvement (e.g., depression, schizophrenia, and the majority manifested a history positive for neuropsychological involvement (e.g., head injury, exposure to neurotoxins, alcoholism). As is generally found with a population as complex as one sampled here, it is frequently not possible to definitively exclude the functionally disordered and to focus exclusively, then upon the cortically impaired. As Malec (1978), Lenzer (1980), and Tucker (1981) have pointed out, it is a tenable proposition that those with major psychological illness manifest cortical damage, or at least disfunction, of some sort.[12] Certainly it has been demonstrated that damage to the cerebral cortex can, and frequently does result in disruptions in the personality's functioning which are for all intent and purpose indiscriminable from their analogs obtaining in the (presumable) absence of cortical insult. All but three of the subjects were males.

Another variable not controlled was that of psychoactive medication. As tends to be characteristic of inpatient samples, the current set of subjects were, by and large, recipients of medications intended to resolve or alleviate undesirable emotional and cognitive consequences or correlates of their functional disorders. These medications included the usual array of antipsychotic and antidepressant agents.

An incisive and extensive review of the literature by Heaton and Crowley (1981) has revealed that once patients have been established on these preparations, performance at neuropsychological measures in not appreciably affected, relative, that is, to unmedicated performance in the same patients. Exceptions to this rule were that tasks demanding focused attention were actually performed better with a stabilized medication than without it, among schizophrenic patients.

In the present sample, testing with schizophrenic patients was deferred, if necessary, until psychotic manifestations had been resolved with a stabilized antipsychotic administration. Patients also were not evaluated in those circumstances in which they were apparently overwhelmed by medication.

Examiners

Neuropsychological evaluations were, in very large part, completed by one of two licensed, Ph.D. level Clinical Psychologists. Both were thoroughly trained in the administration, scoring, and interpretation of the measures used (these are described below, in the following section). A minority of the subjects involved were evaluated by one of two Psychological Technicians, who had been thoroughly trained, and who were carefully supervised, by the senior of the two clinical psychologists involved. A small number (perhaps a dozen) of the subjects were assessed by advanced doctoral graduate students in clinical psychology. These examiners, again, were carefully trained and fully supervised in the administration of the tests. The training of examiners, whether technicians or graduate students, occupied several weeks of full-time study and supervised practice. Examiners were not permitted to evaluate subjects until it was meticulously demonstrated that they adhered to standard techniques of test administration. Over the course of data collection, examiners were also periodically rechecked for "drift" from standard procedures.

The tests administered required, of the administrators a reasonable capacity to adhere to clearly defined instructions, and relatively acceptable manual dexterity, in some cases. Virtually no clinical judgment was required, and interpersonal skills needed were no more sophisticated than those required for a pleasant demeanor. The tests were "objective", in the sense that measured behavior took the form of correct or incorrect responses, times (i.e, latencies) required to respond, and, in some cases, the presence or absence of clinical signs (e.g., dysphasia, dysstereognosis). None of the measures involved were "projective" in character, and no inferences were required in the scoring of the various tests.

Procedures

Subjects were administered, among a broader variety of neuropsychological and more traditionally psychological measures, the complete Halstead-Reitan Neuropsychological Test Battery (HRNTB), with standard (Reitan, 1969) equipment and instructions. The HRNTB is described in detail elsewhere (Reitan and Davison, 1974, pp. 366-370). The HCT, of particular relevance here, is focused upon in the Appendix. The version employed is that which has been in use since the middle 1950's (e.g., Reitan, 1955; Simmel and Counts, 1957; Shore and Halstead, 1958), but not that which Halstead (1943, 1945, 1947, 1951) originally developed. The newer form, as is indicated in the Appendix, was shortened from 360 to 208 items, and from nine to seven subtests. The HCT was administered and scored in strict adherence with Reitan's (1969) instructions, which differ only slightly from those first devised by Halstead.

Design and Analysis

The major concern was with developing an account of HCT item behavior, on the part of the subjects. Specifically, it was undertaken to mathematically define HCT item behavior as a function of the various test-determined factors described above and below, with consideration also given to brain damage. In essence, the focus of analyses was to demonstrate convincingly, and statistically, that learning is a relevant determinant of HCT item response behavior. The analytic designs used were complex, both because the structure of the test itself is complex, and because this structure is systematic enough to permit entering of its characteristics into prediction models. The general model chosen for the several analyses discussed below was multiple linear regression (MLR). The specific MLR orientation which was adopted was that espoused by Cohen and Cohen (1975). The rationale for selecting a generic MLR rather than a generic ANOVA Treatment was based upon the complexity and frequent lack of balance in the designs. The current investigator was aware that the two approaches coincide mathematically, but the computing algorithms available for ANOVA were nowhere near so flexible as those for MLR.

Four separate, though similar, analytic designs were employed during analyses. These involved treating subtests III through VI as separate records of item response behavior. Subtests I and II were not considered as they: (1) were essentially designed to introduce subjects to the HCT format; (2) rarely caused subjects any difficulty: (3) therefore, were of extremely low variances, since nearly all subjects correctly responded to all items. Subtest VII was not included, either, because it appeared to be more a test of memory than of learning.[24] It also lacked the systematic organization apparent in the other subtests.

The factor, 'Trial', figured prominently in each design. Trial refers to the item number within a cluster. For example, subtest III consists of eight clusters of four trials, and eight additional clusters of one trial. As a design factor, Trial is of special importance, in that it represents learning _within_ clusters. The remaining factors have already been considered in detail above.

Analytic designs were those developed for treatment of repeated measures (Cohen and Cohen, 1975, pp. 403-426; Winer, 1971, pp. 514-603) with multiple factors involved.

The dependent measure was item behavior (i.e., correctness or incorrectness in response). With repeated measures designs, between-subjects variance is first computed and partialed away, leaving a composite of systematic and random (i.e., "error") variance, within-subjects. Observations (in this case, _items_) are treated as m separate, distinct examples of the dependent measure, thus, expanding the number of experimental units[25] from n (i.e., number of subjects) to nm. Pertinent research factors were entered analytically by means of dummy-coded variables.

Analyses proceeded according to a strategy which was partly hierarchical, and partly simultaneous. This was because many of the factors and interactions among them were more or less irrelevant to the purpose of demonstrating the operation of learning. However, these "irrelevant" sources of variance were nonetheless felt to be powerful determinants of test behavior. They were "removed" (i.e., partialed) first, in order that learning might be more clearly demonstrated to exist.

Tests of significance of main effects and interactions were F-statistics with appropriate degrees of freedom. Since MLR was used, F-tests were derived from changes in squared multiple correlation estimates. Estimates of error terms used were attributable in large part to Winer (1971, pp. 514-603), although the current investigator disagreed with Winer sometimes, and consequently adopted alternate definitions for error, in certain cases.[26]

Although each of the four subtests involved was treated separately, the same "generic" design and analytic strategy was applied to all. This is summarized, in condensed form, in tables 1, 2, 3, and 4. The ensuing discussion follows from the information contained in these four tables.

Note that table 1 lists 22 effects, or components of variance. Tables 2, 3, and 4, however, divide this array of effects into more meaningful sets.

The component, 'subjects' listed in table 1, refers to variance attributable to differences between subjects. This variance was important both because it required partialing, or removal, from the within-subjects or replicated part of the design, and because an attempt was made to account for part of this variance by entering the factor, 'damage'. 'Damage' was defined as the augmented Impairment Index. The HCT was removed from the array of indicators ordinarily utilized in computing the Impairment Index, and the Trail

Table 1. Summary of Effects Considered.

Effect	Level of Measurement
(1) Subjects	Nominal[a]
(2) Damage	Interval[b]
(3) Option	Nominal
(4) Sequence	Nominal
(5) Trial	Nominal
(6) Sequence power one[c]	Ordinal
(7) Trial power one [c]	Ordinal
(8) Sequence power two[d]	Ordinal
(9) Trial power two[d]	Ordinal
(10) Sequence power three[e]	Ordinal
(11) Trial power three[e]	Ordinal
(12) Option X Sequence	Nominal
(13) Option X Trial	Nominal
(14) Trial X Sequence	Nominal
(15) Damage	Interval
(16) Sequence X Damage	Interval
(17) Trial X Damage	Interval
(18) Option X Damage	Interval
(19) Trial X Sequence X Option	Nominal
(20) Trial X Sequence X Damage	Interval
(21) Sequence X Option X Damage	Interval
(22) Trial X Option X Damage	Interval

[a]The term, 'nominal,' may be reasonably supplanted by 'qualitative'.

[b]Likewise, 'interval' may here be understood as roughly synonomous with 'quantitative'.

[c]Coded by means of power 1 orthogonal polynomials.

[d]Coded by means of power 2 orthogonal polynomials.

[e]Coded by means of power 3 orthogonal polynomials.

Table 2. Between-Subjects Analytic Model

	Effect	Degrees of Freedom[a]	Denominator[b]
(1)	$R^2y.s$	n-1	none
(2)	$R^2y.d$	1/n-1	(1)[c]-(2)

Note: The notation used in this table is largely consistent with that Cohen and Cohen (1975) favor, and similar as well to that espoused by many others. Thus:

R^2 = a (any) squared multiple correlation;

$R^2y.s$ = the proportion of variance in the dependent measure attributable to differences between subjects.

The remaining "Effect" quantity is interpreted as the proportion of variance attributable to, or accounted for, or by, the various independent measures or arrays of same. The parenthesized numbers occurring to the 68t of the "Effect" quantities refer back to Table 1, and consequently ought to reduce confusion.

[a]The convention, df (numerator/df (denominator) was adopted and adhered to, throughout.

[b]"Denominator" may be taken to indicate the "error term," or the sum of squares eventually entering the F-ratio as the denominator.

[c]As implied by the "Note," above, the parenthesized numerals are actually abbreviations for their associated effects. Thus:

$$(1) = R^2y.s;$$

$$(1) - (2) = R^2y.s - R^2y.d.$$

Table 3. Basic Within-Subjects Analytic Model

	Effect	Degrees of Freedom[a]	Denominator[b]
(3)	R^2y.o	Co-1/n[c](m-1)-(Co^{-1})	1-(1+3)
(4)	R^2y.se	Cse-1/D(3)-(Cse^{-1})	1-(S(3)+4)
(5)	R^2y.t	Ct-1/D(4)-(Ct^{-1})	1-(S(4)+5)
(12)	R^2y.oXse	CoXse-1/D(5)-(CoXse^{-1})	1-(S(5)+12)
(13)	R^2y.oXt	CoXt-1/D(12)-(CoXt^{-1})	1-(S(12)+13)
(14)	R^2y.tXse	CtXse-1/D(13)-(CtXse^{-1})	1-(S(13)+14)
(15)	R^2y.d	1/D(14)$^{-1}$	1-(S(14)+15)
(16)	R^2y.seXd	Cse-1/D(15)-(Cse^{-1})	1-(S(15)+16)
(17)	R^2y.tXd	Ct-1/D(16)-(Ct^{-1})	1-(S(16)+17)
(18)	R^2y.oXd	Co-1/D(17)-(Co^{-1})	1-(S(17)+18)
(19)	R^2y.tXseXo	CtXseXo-1/D(18)-(CtXseXo^{-1})	1-(S(18)+19)
(20)	R^2y.tXseXd	CtXse-1/D(19)-(CtXse^{-1})	1-(S(19)+20)
(21)	R^2y.seXoXd	CseXo-1/D(20)-(CseXo^{-1})	1-(S(20)+21)
(22)	R^2y.tXoXd	CtXoXd-1/D(21)-(CtXoXd^{-1})	1-(S(21)+22)

Note. Refer to the general note applying to Table 2. It is germane here, as well.

[a]The subscripted symbols, C_i, indicate the number of levels for a given independent variable, or "condition," as Cohen and Cohen (1975, e.g., pp. 403-426) might say. The entries, D(i), appearing in the denominators' degrees of freedom, represent the denominators for the ith, usually immediately preceding, steps.

[b]Parenthesized numerals refer, as was true of Tables 1 and 2, to designated effects variance components. The entries, S(i) represent the entire parenthesized quantities from the immediately preceding steps. Thus:

S(4) = (S(3) + 4) = (1+3+4).

[c]In this statement, 'n' represents the number of subjects, usually 159; 'm' represents the number of items in a subtest, e.g., 32 for subtest III.

Table 4. Within-Subjects Analytic Model for Polynomial Components.

	Effect	Degrees of Freedom	Denominator
(3)	$R^2y.o$	$Co-1/n(m-1)-(Co-1)$	$1-(1+3)$
(6)	$R^2y.se^1$	$Cse^1-1/D(3)-(Cse^1-1)$	$1-(S(3)+6)$
(7)	$R^2y.t^1$	$Ct^1-1/D(6)-(Ct^1-1$	$1-(S(6)+7)$
(8)	$R^1y.se^2$	$Cse^2-1/D(7)-(Cse^2-1)$	$1-(S(7)+8)$
(9)	$R^2y.t^2$	$Ct^2-1/D(8)-(Ct^2-1)$	$1-(S(8)+9)$
(10)	$R^2y.se^3$	$Cse^3-1/D(9)-(Cse^3-1)$	$1-(S(9)+10)$
(11)	$R^2y.t^3$	$Ct^3-1/D(10)-(Ct^3-1)$	$1-(S(10)+11)$

<u>Note</u>. The general note associated with Table 2 applies here, as well. Other notes appearing in Table 3 are applicable here. Finally, the suprascripts refer to the degrees of polynomials involved in the effects. All are orthogonal.

Making Test, Parts A and B, were added.[27] Standard (Reitan, 1969) cutoff points were utilized, and each indicator was given a score of 1 (presumptive brain damage) or 0 Insert (presumptive absence of brain damage). The 8 scores involved were summed to produce the augmented Impairment Index. In this manner, the HCT was not used in predicting components of itself, and the loss in discriminating power of the Impairment Index, effected by removal of the HCT, was partially offset by augmenting the quantity with Trails A and B.

The factor, 'option', refers to the item response which is the correct one. 'Sequence' refers to the ordinal position of clusters of items in each of the subtests. 'Trial' indicates the order of the items within sequences, or clusters.[28]

Consistent with the hypotheses as spelled out above, the following predictions were expected to obtain. Damage was expected to predict HCT errors. Fundamentally, this was the passe expectation that HCT subtest behavior would correlate with other indices of brain damage. The factor was entered largely because it proved critical to within-subjects comparisons. It also was appropriate to consider this factor with the object of accounting for as much variance as possible, between and within subjects.

Option was expected to account for a considerable proportion of within-subjects variance. Relatively little attempt was made to interpret this effect, for two reasons. First, the principal rationale for having entered it at all was that in general this factor accounts for significant systematic variance which would otherwise be included as "error" in the comparisons concerning learning, as outlined below. The second reason was that Simmel and Counts (1957) have already spent considerable energy covering the topic of the impact option can be expected to have upon HCT item behavior. The interested reader is encouraged to consult this source. For purposes herein, option can be conceptualized as item stimulus characteristics which induce subjects to make some types of errors more frequently than others.

Sequence was predicted to attain significance. This factor served as a measure of learning, in that it was predicted that item errors would decrease as a function of time and experience for the subject, or as a function of sequence, itself, when conceptualized ordinally. No predictions were made concerning the individual levels of the factor, 'Sequence'.

Trial was a second component of learning. Specifically, and as was true with sequence, it was expected that learning would obtain as a function of trial. It was expected to emerge as significant.

Both sequence and trial were also considered by means of orthogonal polynomials of first, second, and third degree. The summary of this part of the analyses appears in table 3. These comparisons were elected, because although it was expected that learning could adequately be represented linearly, the possibility was felt to exist that a learning function of an exponential and asymptotic character might obtain, approximating a third degree statement. It was predicted that second and third degree components would prove significant for both sequence and trial, although the prediction was tendered with more faith for sequence than trial, because the former variety of learning was permitted by the design to continue for a longer time.

Interaction (12) was included so as to evaluate the hypothesis that option decreases in salience over time. This effect implied that learning was occurring. The same was true for interaction (13) although again, that prediction was held with less faith than the one concerning interaction (12). Interaction (14) implied that learning within sequences might alter as a function of time. Indeed, it was predicted that this interaction would attain significance, and that learning vis-a-vis trials would accelerate, over time.

Component (15) was included so as to partial the impact of brain damage by itself from regression equation prior to entering interactions involving this variable. In other regards, it was entirely irrelevant.

Interaction (16), (17), and (18) all were concerned with the issue of whether brain damage would exert a qualitative rather than a quantitative impact upon HCT item performance characteristics. As such, tests of those effects could be brought to bear on Reitan's (1958, 1959) conclusions that brain damage would affect reasoning in a quantitative, rather than in a qualitative way. Specifically, significance of one or more of those interactions would provide a basis for inferring that the distinction between brain damaged and non-brain damaged reasoning was a qualitative matter. Based upon Reitan's (1958, 1959) findings, these interactions were predicted to fail.

Interaction (19) implied that the impact of Option upon responding, within sequences or across trials would decrease over time. This, then, was an additional aspect of learning. It was expected to emerge as significant.

Interactions (20), (21), and (22) were also concerned with the impact of brain damage upon the learning process. Component (20) pertained to the impact brain damage exerted upon the acceleration of learning over time. Component (21) referred to the effect brain damage might have upon reduction in importance of option, over time. Component (22) was concerned with the consequence brain damage implied insofar as the reduction of importance of option across trials was concerned. All three of those interactions were expected to fail, in keeping with Reitan's (1958, 1959) accounts.

Because items 33-40 in subtest III of the HCT lacked the organized systematicity inherent in the first 32 items, these last 8 items were dropped from consideration in the analyses.

For subtest IV, all 40 items were involved, and these produced 10 sequences of from 3 to 6 trials. Obviously the design was unbalanced, requiring adjustments in degrees of freedom.

Subtest V involved 40 items organized into 6 sequences of from 3 to 9 trials.

Subtest VI was unusual in three ways. First, it utilized the same "organizing principle", or "set" as subtest V: quadrants numbered clockwise from the upper left, 1 through 4. Second, the first six items were repli-cations of stimuli found in subtest V. Finally the last 10 items were explicitly unrelated to the fist 30, or more pertinently, to the middle 24 (consult the Appendix for specifics). In order to resolve the logistics problems posed by these eccentricities, it was decided to truncate this subtest at either end, and employ only items 7 through 30. This decision produced 24 items organized into 3 sequences of 8 items each. Like subtest III, this design was balanced in terms of trial and sequence.

Again, essentially identical factors were considered and essentially identical designs were implied in the cases of all 4 of the subtests involved. Relevant changes from subtest to subtest involved only alterations in the number of levels of the factors, 'Trial' and 'Sequence'. This shifted degrees of freedom to some extent, both in the individual effects, and in the interactions involved.

Results

Age, education, Augmented Impairment Index, and summary scores for each of the measures figuring into the Augmented Impairment Index were computed for each subject. These quantities appear in Table 5.

Table 5. Summary of Sample Characteristics and Performance at Relevant Variables.

Variable	n	Mean	S.D.
Age	141	37.93	12.49
Education	136	10.98	2.28
Factual Performance test (Total Time)	159	22.51	9.97
Factual Performance test (Memory)	159	6.40	2.07
Factual Performance test (Localization)	159	3.02	2.34
Seashore Rhythm test	159	6.07	4.17
Speech Perception test	159	10.89	6.90
Finger Oscillation (Dominant flood)	159	46.14	8.54
Trail Making test (Part A)	159	46.11	27.02
Trail Making test (Part B)	159	133.15	93.91
Augmented Impairment Index	159	4.74	2.18

Age and education were computed in terms of decimal years. For education, this was completely straightforward. For age, it was necessary to first transform days and months into their decimal equivalents. The values for Tactual Performance Test total time were computed in decimal minutes. The localization and memory components were based upon blocks drawn correctly and/or in their correct positions. Seashore Rhythm Test was phrased as the rank score in preference to the number of erroneous or correct responses. Speech Perception Test was phrased in terms of total errors made. Finger Oscillation was based upon the mean number of oscillations performed during a 10 second interval. Both parts A and B of the Trail Making Test were defined as decimal seconds required to complete the task. The Augmented Impairment Index was defined as the number within the impaired range, in accordance with Reitan's (1969) empirically derived cutoff values. As can be seen, the mean was something over one half of the indicants filling within the impaired range, for this sample.

In order to develop tests of the hypotheses, it was first essential to partition the variance of item responses into between-subjects and within-subjects components. Though there exist several approaches to resolving this problem, the procedure of Cohen and Cohen (1975, pp. 403-428) was adopted. Thus, the mean item response for each subject, within subtests was computed, and the mean and variance of these means were taken. The variance of these means amounted to the variance between-subjects. Then, within each subtest, the mean and variance was computed across subjects, and with the individual item defined as the unit of analysis. This variance estimate amounted to the total variance. By dividing the variance in subjects' mean item responses by the variance across items, it was possible to derive estimates of within-subjects, and between-subjets variance, relative to total variance, for each of the subtests. These quantities are summarized in Tables 6 and 7.

Following the portioning of variance into between-subjects and within-subjects components, the regression of brain damage, as defined by the augmented Impairment Index, upon subjects' mean item scores within subtests, was carried out. This amounted to the between-subjects portion of the analyses, and results, by subtest, are found in Table 8.

In table 8, the column designated "R" contains the estimates of multiple correlation between the augmented Impairment Index and the subjects' mean item response, for each subtest. The next column, designated R^2, is, of course, the squared multiple correlation, and may be directly interpreted as the portion, or proportion, of variance in subjects' mean item responses attributable to their functional status, or defined by the Augmented Impairment Index. The next column contains degrees of freedom, and the convention, numerator/denominator, was adopted. Following this, the F-ratio and its significance appear. Significance estimates for subtests IV, V, and VI are not precise, though in each case the F-ratio vastly exceeds .001 in significance, by an indeterminate degree. Power analyses proceeded in the

Table 6. Summary of Subject and Item Means, Variances, and Standard Deviations, by Subtest.

Subtest	Subject				Item (Total)			
	n	Mean[a]	Variance	S.d.	NM[b]	Mean	Variance	S.D.
III	156	.535	.065	.254	5966	.535	.249	.499
IV	155	.437	.080	.284	6053	.432	.245	.495
V	150	.428	.029	.170	5942	.425	.244	.494
VI	149	.230	.043	.206	3416	.223	.173	.416

[a]Means indicate average errors, rather than average correct responses. To obtain p-values, in the usual sense, it would be necessary to derive the complements of these values, by subtracting them from unity.

[b]The use of 'nm' is meant to indicate that the total number of observations, or units of analysis, is the product of the number of subjects, 'n', and the number of items, 'm', for a given subtest.

Table 7. Derivation of Between-Subjects and Within-Subjects Estimates of Variance

Subtest	Subjects' Variance	Total Variance	Between-Subjects Variance	Within-Subjects Variance
III	.065	.249	.260	.740
IV	.080	.245	.328	.672
V	.029	.244	.118	.883
VI	.043	.173	.245	.755

Table 8. Between-Subjects Variance Analysis, or Regression of Brain Damage Upon Average HCT Item Response, Organized By Subtest.

Subtest	R	R^2	D.F.	F[a]	Significance	Power[b]
III	.225	.051	1/154	8.241	.055	.882
IV	.520	.271	1/153	56.718	.001	.999
V	.499	.249	1/148	49.098	.001	.999
VI	.447	.200	1/147	36.683	.001	.999

[a]For all subtests, F was derived as follows:

$$F = \frac{R^2}{1-R^2} \times \frac{D.F.(Den.)}{D.F.(Num.)}$$

[b]Based upon winer's (1971) algorithms utilizing the noncentral t-distribution, with alpha set at .05, and significance testing being one-tailed. Quantitives estimated at .999 are actually greater than this, albeit by an indeterminate margin.

customary manner, and by any reasonable standard, the F-tests obtaining manifested completely acceptable levels of power.

Though the proportion of variance attributable to brain damage shifted somewhat among the subtests, from just over 5% (Subtest III) to slightly more than 27% (subtest IV), it seemed clear that cognitive impairment, as established with the augmented Impairment Index, was a valid predictor of performance at HCT subtests III through VI, as they have been defined here. This was not particularly surprising, and can only be viewed as additional replications of the work summarized in the introductory sections of this monograph.

From the quantities appearing in Table 7, it also was noted that the proportion of variance attributable to differences within-subjects ranged from a low of about 17% (subtest IV) to a high of 88% (subtest V), with the remaining two values at approximately 75% (subtests III and VI). By a substantial margin, then, it was inferred that item characteristics determined or produced the majority of the variance in responses for each of the subtests. This was in fundamental agreement with the findings of Simmel and Counts (1957), who found that the qualities of the HCT items were extremely influential determinants of subjects' behavior, whether they were brain damaged or not.

Tables 9 through 16 summarize the basic and polynomial within-subjects regression analyses for the four subtests. The tables are arranged in pairs, with the basic analysis

Table 9. Subtest III: Basic Within-Subjects Regression Analysis.

Effect	Unique R^2	Cumulative R2	D.F.	F	Significance	Power[b]
Option	.010	.010	3/4807	16.463	.001	.420
Sequence	.045	.055	7/4800	32.522	.001	.990
Trial	.013	.068	3/4797	22.530	.001	.557
Option X Sequence	.021	.089	12/4785	9.292	.005	.990
Option X Trial	.002	.092	5/4780	2.315	.05	.152
Trial X Sequence	.001	.092	1/4779	4.476	.05	.100
Damage	.014	.107	7/4778	75.888	.001	.349
Sequence X Damage	.002	.108	7/4771	1.169	n.s.	.136
Trial X Damage	.004	.112	3/4768	7.394	.01	.210
Option X Damage	.001	.113	3/4765	.949	n.s.	.100
Trial X Sequence X Option	0	.113	0/4767	--	n.s.	.100
Trial X Sequence X Damage	.001	.113	21/4744	.150	n.s.	.100
Sequence X Option X Damage	.0001	.113	5/4739	.054	n.s.	.100
Trial X Option X Damage	.00001	.113	9/4737	.027	n.s.	.100

Note. Estimate of unique and cumulative R^2 are based upon, or phrased relative to, within-subjects variance alone.

[a]Significance estimates were frequently imprecise. Results were significant at or beyond the level specified, but for the level, .001, significance may exceed the value by an indeterminate amount.

[b]Values of .990 are actually in excess of this by an indeterminate amount; values of .100 are actually less (poorer) than this by an indeterminate amount.

Table 10. Subtest III: Polynomial Within-Subjects Regression Analysis

Effect	Unique R^2	Cumulative R2	D.F.	F	Significance	Power[b]
Option	.010	.010	3/4807	16.463	.001	.420
Sequence (Linear Component)	.034	.044	1/4806	170.643	.001	.648
Trial (Linear Component)	.005	.049	1/4805	23.387	.001	.139
Sequence (Quadratic Component)	.001	.050	1/4804	6.372	.05	.100
Trial (Quadratic Component)	.009	.059	1/4803	44.032	.001	.222
Sequence (Cubic Component)	.008	.067	1/4802	42.768	.001	.217
Trial (Cubic Component)	0	.067	1/4801	0	n.s.	.100

[a]Significance estimates were relatively imprecise. Results were significant at or beyond the level specified, but for .001, significance may exceed the stated value by an unknown margin.

[b]Values of .100 are actually less (poorer) than this by an unspecified margin.

Table 11. Subtest IV: Basic Within-Subjects Regression Analysis

Effect	Unique R^2	Cumulative R^2	D.F.	F.	Significance[a]	Power[b]
Option	.042	.042	3/5895	87.155	.001	.971
Sequence	.019	.061	9/5886	13.005	.001	.971
Trial	.009	.070	5/5881	11.512	.005	.534
Option X Sequence	.016	.086	19/5862	5.334	.001	.990
Option X Trail	.001	.087	3/5859	1.454	n.s.	.100
Trial X Sequence	0	.087	0/5859	---	n.s.	.100
Damage	.134	.220	1/5858	1003.958	.001	.990
Sequence X Damage	.001	.221	9/5849	.684	n.s.	.103
Trial X Damage	.0005	.222	5/5844	.676	n.s.	.100
Option X Damage	.001	.222	3/5841	1.552	n.s.	.100
Trial X Sequence X Option	.007	.229	28/5813	1.874	.05	.967
Trial X Sequence X Damage	.004	.233	22/5791	1.287	n.s.	.644
Sequence X Option X Damage	0	.233	23/5768	0	n.s.	.100
Trial X Option X Damage	0	.233	7/5761	0	n.s.	.100

<u>Note</u>. Estimates of unique and cumulative R^2 are based upon, or phrased relative to, within-subjects variance alone.

Table 12. Subtest IV: Polynomial Within-Subjects Regression Analysis

Effect	Unique R^2	Cumulative R^2	D.F.	F.	Significance[a]	Power[b]
Option	.042	.042	3/5895	87.155	.001	.971
Sequence (Linear Component)	.001	.044	1/5894	6.532	.05	.100
Trial (Linear Component)	.005	.049	1/5893	30.781	.001	.146
Sequence (Quadratic Component)	.0001	.049	1/5892	.434	n.s.	.100
Trial (Quadratic Component)	.0002	.049	1/5891	1.301	n.s.	.100
Sequence (Cubic Component)	.001	.050	1/5890	7.812	.01	.100
Trial (Cubic Component)	.003	.053	1/5889	18.594	.005	.107

[a]Significance estimates often were imprecise. Per the level, .001, significance may well exceed the stated value hy an unknown amount.

[b]Values of .100 are actually less (poorer) than this, by an unknown amount.

Table 13. Subtest V: Basic Within-Subjects Regression Analysis

Effect	Unique R^2	Cumulative R^2	D.F.	F.	Significance[a]	Power[b]
Option	.015	.015	3/5789	28.690	.001	.564
Sequence	.136	.151	5/5784	185.593	.001	.990
Trial	.017	.168	8/5776	14.892	.001	.960
Option X Sequence	.040	.208	14/5762	20.635	.001	.990
Option X Trail	.007	.215	7/5755	7.773	.001	.605
Trial X Sequence	.003	.218	2/5753	10.816	.001	.146
Damage	.031	.249	1/5752	234.282	.001	.692
Sequence X Damage	.006	.255	5/5747	9.440	.001	.439
Trial X Damage	.001	.256	8/5739	.964	n.s.	.115
Option X Damage	.0001	.256	3/5736	.565	n.s.	.100
Trial X Sequence X Option	.003	.259	26/5710	.857	n.s.	.554
Trial X Sequence X Damage	.010	.269	29/5681	2.677	.005	.990
Sequence X Option X Damage	0	.269	0/5681	---	n.s.	.100
Trial X Option X Damage	0	.269	19/5662	0	n.s.	.100

<u>Note</u>. Estimates of unique and cumulative R^2 are based upon, or phrased relative to, within-subjects variance alone.

Table 14. Subtest V: Polynomial Within-Subjects Regression Analysis

Effect	Unique R^2	Cumulative R^2	D.F.	F.	Significance[a]	Power[b]
Option	.015	.015	3/5789	28.690	.001	.564
Sequence (Linear Component)	.005	.020	1/5788	28.927	.001	.139
Trial (Linear Component)	.025	.045	1/5787	153.162	.001	.510
Sequence (Quadratic Component)	.018	.063	1/5786	109.432	.001	.385
Trial (Quadratic Component)	.0002	.063	1/5785	.988	n.s.	.100
Sequence (Cubic Component)	.003	.066	1/5784	21.307	.005	.115
Trial (Cubic Component)	.003	.069	1/5783	17.454	.005	.102

[a]Significance estimates frequently were imprecise. Per alpha of .001, results obtaining generally exceeded this level, though by an unknown extent.

[b]Values of .100 were actually less (poorer) than this, albeit to an unknown extent.

Table 15. Subtest VI: Basic Within-Subjects Regression Analysis

Effect	Unique R^2	Cumulative R^2	D.F.	F.	Significance[a]	Power[b]
Option	.020	.020	3/3264	21.672	.001	.703
Sequence	.137	.156	2/3262	264.802	.001	.990
Trial	.012	.168	7/3255	6.716	.001	.818
Option X Sequence	.012	.180	6/3249	7.621	.001	.758
Option X Trail	.003	.183	5/3244	2.327	.05	.205
Trial X Sequence	0	.183	0/3244	---	n.s.	.100
Damage	.056	.239	1/3243	238.037	.001	.907
Sequence X Damage	.008	.247	2/3241	17.278	.001	.335
Trial X Damage	.001	.248	7/3234	.645	n.s.	.112
Option X Damage	.001	.249	3/3281	---	n.s.	.100
Trial X Sequence X Option	0	.249	0/3231	---	n.s.	.100
Trial X Sequence X Damage	.002	.251	11/3220	.766	n.s.	.226
Sequence X Option X Damage	0	.251	0/3220	---	n.s.	.100
Trial X Option X Damage	0	.251	9/3211	0	n.s.	.100

Note. Estimates of unique and cumulative R^2 are phrased relative to only within-subjects variance.

Table 16. Subtest VI: Polynomial Within-Subjects Regression Analysis

Effect	Unique R^2	Cumulative R^2	D.F.	F.	Significance[a]	Power[b]
Option	.020	.020	3/3264	21.672	.001	.703
Sequence (Linear Component)	.100	.119	1/3263	369.896	.001	.978
Trial (Linear Component)	.004	.123	1/3262	14.436	.001	.128
Sequence (Quadratic Component)	.037	.160	1/3261	143.478	.001	.722
Trial (Quadratic Component)	.001	.161	1/3260	3.070	n.s.	.100
Trial (Cubic Component)	.004	.164	1/3259	13.221	.001	.121

[a]Significance estimates frequently were imprecise. Those appearing as .001 were actually somewhat in excess of this, but to an unknown extent.
[b]Power estimates of .100 were actually less (poorer) than this, although to an inexplicit extent.

being presented first, before the polynomial analysis. The power of each F-test computed was also estimated, and resulting quantities were included in these tables.

Power has been defined as the complement of beta, or the probability of failing to reject the null, when it is indeed false. Power, then, is the probability of rejecting the null, when it is false. The quantity, power, increases as a function of sample size, effect size, and the size of alpha. Specifically, other things being held constant, as the sample size is increased, power will increase, as the effect size (or the difference between the populations) increases, power will increase, as alpha is set more liberally, power will increase.

The purpose or objective of high or large power is to assure that the null will be rejected when it "ought" to be, when it is in fact false. Cohen and Cohen (1975, pp. 117-118) recommend selecting as appropriate some value between .70 and .90, say, .80 as the lowest acceptable value for power. Once this has been fixed, then, the investigator is in the position of manipulating the other three parameters, but chiefly, effect and sample sizes, in her or his design. Alpha is generally set at .05 for reasons involving the relative costs of accepting or rejecting the null hypothesis when it is true.

In applying these considerations to the present set of within-subjects variance or regression analyses, the issue of power emerges as marginally irrelevant, for the reason that sample sizes are virtually infinitely large, rendering it highly unlikely that genuine departures from the null hypothesis, however slight, will go undetected, given a reasonable alpha of .01 or .05. In fact, alphas of .001 or even .0001 are fairly liberal, given the sample sizes involved. Under this circumstance, power approaches meaninglessness, because it stands as an indicant only of effect size. And, as can be readily seen in Tables 9 through 16, those analyses associated with diminishingly small effects, or R^2 estimates are associated, as well, with reduced power of the F-tests. However, sample sizes were so large that the resulting F values tended to be gigantic, and consequently grossly significant, anyhow. Another means of expressing the same idea is that with alpha at .001 or smaller, and a significant F-test, power is unimportant, because the null has already been rather convincingly rejected. For these reasons, although power was routinely computed with alpha fixed at .05, its interpretation was somewhat unimportant.

Overall, the results rather clearly indicated that the variables under consideration accounted for at the most a moderate portion of the variance in subjects' item responses. Thus, for the basic analyses, total variance attributable to the full array of the independent variables was only .113 for subtest III, .233 for subtest IV, .269 for subtest V, and .251 for subtest VI. The orthogonal polynomial components accounted for .057 in subtest III, .011 in subtest IV, .054 in subtest V, and .144 in subtest VI. Though statistical tests were highly significant in many instances, and were carefully evaluated and interpreted, it nonetheless was obvious that the substantial majority of the variance in within-subjects item behavior could not be accounted for on the basis of the factors selected for analytic consideration herein.

Option

The factor 'Option', proved dramatically significant for each subtest, and perhaps especially so for subtest IV, where it alone accounted for more than 4% of the total within-subjects variance. As was stated above, this result was anticipated, and an account of it has already been carefully detailed by Simmel and Counts (1957). While the finding was not of special or particular interest to the present investigator, the item means and variances, segregated by option, were examined. These quantities appear in table 17.

With one exception, partialing away between-subjects variance had no effect upon relative item difficulties, within subtests. The exception was that for subtest V, items 2 and 3 reversed position after the between-subjects variance had been statistically extracted. The reason for this shift remained unclear.

For subtest III, option 4 items proved the most difficult, followed in order of descending difficulty by options 2, 1, and finally, 3. As Simmel and Counts (1957, pp. 27-50) point out, this is consistent with a bias in subjects' response sets toward selecting options 1 and 3 in

Table 17. Item means and Variances, Segregated by Option

Derived With Raw Item Data

Subtest	Option 1		Option 2		Option 3		Option 4	
	Mean	Var.	Mean	Var.	Mean	Var.	Mean	Var.
III	.541	.248	.548	.248	.460	.248	.583	.243
IV	.317	.217	.384	.237	.517	.250	.506	.250
V	.519	.250	.371	.233	.357	.234	.418	.243
VI	.289	.205	.194	.156	.258	.192	.159	.134

Derived with Between-Subjects' Variance Partialed

Subtest	Option 1		Option 2		Option 3		Option 4	
	Mean	Var.	Mean	Var.	Mean	Var.	Mean	Var.
III	.005	.207	.012	.164	-.076	.201	.048	.156
IV	-.114	.158	-.048	.135	.085	.197	.074	.146
V	.093	.225	-.054	.201	-.031	.216	-.008	.211
VI	.063	.164	-.029	.121	.035	.138	-.063	.107

Note. Means derived with raw data are tantamount to conditional probabilities of making an error, rather than a correct response, as is more conventional.

preference to options 2 and 4. As can be seen in Figure I-C in the Appendix, subtest III items are all characterized by three similar and one distinct stimuli. Subjects tend to count, and to select option 3 on the basis of the three similar stimuli, or option 1 on the basis of the single unique stimulus in each item. This bias produces an enhanced likelihood of earning a correct response (although for the "wrong" reasons) for items in which options 1 or 3 are actually correct. From Table 9, it may be noted that this finding is not especially remarkable, as the factor, 'Option', accounts for merely 1% of the within-subjects variance. Power of the F-test was also unacceptably low, at .420, and this value is largely attributable to the small effect size.

For subtest IV, 'Option' accounted for 4.2% of within-subjects variance, and the highly significant F-test was of power, .971. For this subtest, items with option 3 being correct proved the most difficult, followed ordinarily by options 4, 2, and 1. As Simmel and Counts (1957, pp. 51-71) have pointed out, this likely obtained because subjects mistakenly assume that quadrants 3 and 4 in the item stimuli will occur in a left-right position, rather than a right-left, or "clockwise" position, relative to one another. Options 1 and 2 do, in fact, occur in a left-right order, and consequently, items characterized by these correct options have a relatively high frequency of correct responses, as compared to items defined by correct options 3 and 4.

In subtest V, 'Option' accounted for 1.5% of within-subjects variance, and though significant, the effect was small enough to render power unacceptably low, at .564. For this subtest, items having option 1 correct were the most difficult, followed by options 4, 2, and 3. Though for this subtest, the determinants of the nonrandom error distribution across options are somewhat complex, Simmel and Counts (1957, pp. 72-96) point out that items characterized by options 2 and 3 tend to be more reinforcing of the appropriate organizing principle, while items described by options 1 and 4 tend to support or suggest erroneous hypotheses on the part of subjects. In particular, for much of the subtest, option 1 items reinforce the response set which had been learned from subtest IV, and option 4 items suggest a single, unified whole stimulus, biasing subjects in favor of option 1. Appealing to Figure I-E in the Appendix supports these contentions.

The factor, 'Option', accounted for 2% of the within-subjects variance in subtest VI, a significant finding, the F-test of which was associated with power of .703. This estimate of power is somewhat low, and again, the reason for this is the rather minute magnitude of the effect. This subtest proved the easiest for subjects to master as can be readily deduced from the values in Table 17. Essentially, this was because subtests V and VI share the same organizing principle, or accurate response set, and consequently, subtest VI performance can be viewed as practiced, or profiting from prior experience with subtest V. Maximally difficult items were those having option 1 as correct followed by options 3, 2, and finally, 4. According to Simmel and Counts (1957, pp. 97-116), the emergence of option 1 as the more difficult occurred for the same reason as was posited for subtest V. The remaining options can be seen from Table 17 to be characterized by rather low and nearly equal frequences of error, and it was felt reasonable to argue that most subjects had correctly grasped the organizing principle by the time they began subtest VI. There was a slight tendency to favor option 4, and this in all likelihood was because the majority of the items analyzed in this subtest could be readily construed as stimuli consisting of four parts.

The analysis of the salience of 'Option' in determining HCT item behavior, within-subjects was, again, of little importance to the current investigator. The component was built into the design in order to permit its influence to be partialed away in operation for the analyses which <u>were</u> of interest.

<u>Sequence</u>

This effect proved dramatically significant for each of the subtests, and of the effects which were significant, sequence tended to account for more variance than most, hence the rather large estimates for power (more than .95, in all cases). Table 18 contains the means for items segregated by sequence, for each of the subtests were devised on the bases both of raw item data, and as well, after having

Table 18. Marginal Means for Sequence

Derived With Raw Item Data

Subtest	Sequence									
	1	2	3	4	5	6	7	8	9	10
III	.711	.639	.472	.517	.530	.549	.452	.410	--	--
IV	.397	.472	.591	.430	.405	.334	.450	.366	.492	.430
V	.571	.356	.479	.151	.717	.361	--	--	--	--
VI	.417	.124	.121	--	--	--	--	--	--	--

Derived With Partialed Data

Subtest	Sequence									
	1	2	3	4	5	6	7	8	9	10
III	.179	.085	-.030	-.027	.000	-.006	-.062	-.142	--	--
IV	-.033	.066	.131	-.007	-.029	-.070	-.031	-.011	.065	.040
V	.134	.074	.056	-.272	.293	-.031	--	--	--	--
VI	.180	-.099	-.088	--	--	--	--	--	--	--

Note. Both between-subjects variance and that variance attributable to correct option was partialed from raw item data in preparing the second half of the tabled values. Tabled valves, consequently, may be thought of as standard scores, rather than probabilities of erroneous responses.

partialed away variance attributable to between-subjects differences and the factor, 'Option'.

The predicted relationship between sequence, or, more lucidly, familiarity or experience with a given organizing principle, and probability of errors at HCT items, was not unequivocally observed across HCT subtests. More lucidly phrased, sequence did indeed exert an impact upon the probability of making errors in response to HCT item stimuli, but the effect of sequence upon this phenomenon was dramatically more complex than had been anticipated. As much can be readily deduced from even a cursory glance at Figures 1 through 4, in which the partialed sequence means for the four relevant HCT subtests were graphed.

The curve depicted in Figure 1 was basically consistent with what was predicted, in that its slope, to a linear approximation, was negative. The linear component for sequence accounted for 3.4% of the within-subjects item variance, and this amount proved highly significant.

However, the linear form of the curve was somewhat disfigured by a perturbation occurring at sequences 5 and 6, where apparently items become more difficult. From examining the items of this subtest portrayed in Figure I-C, it was determined that this coincided with the strings of items 17 through 20 (sequence 5) and 21 through 24 (sequence 6). These strings of items are characterized by the introduction of a "distractor" feature, as it were, in that suddenly individual stimuli comprising items differ in two characteristics, only one of which is germane to the organizing

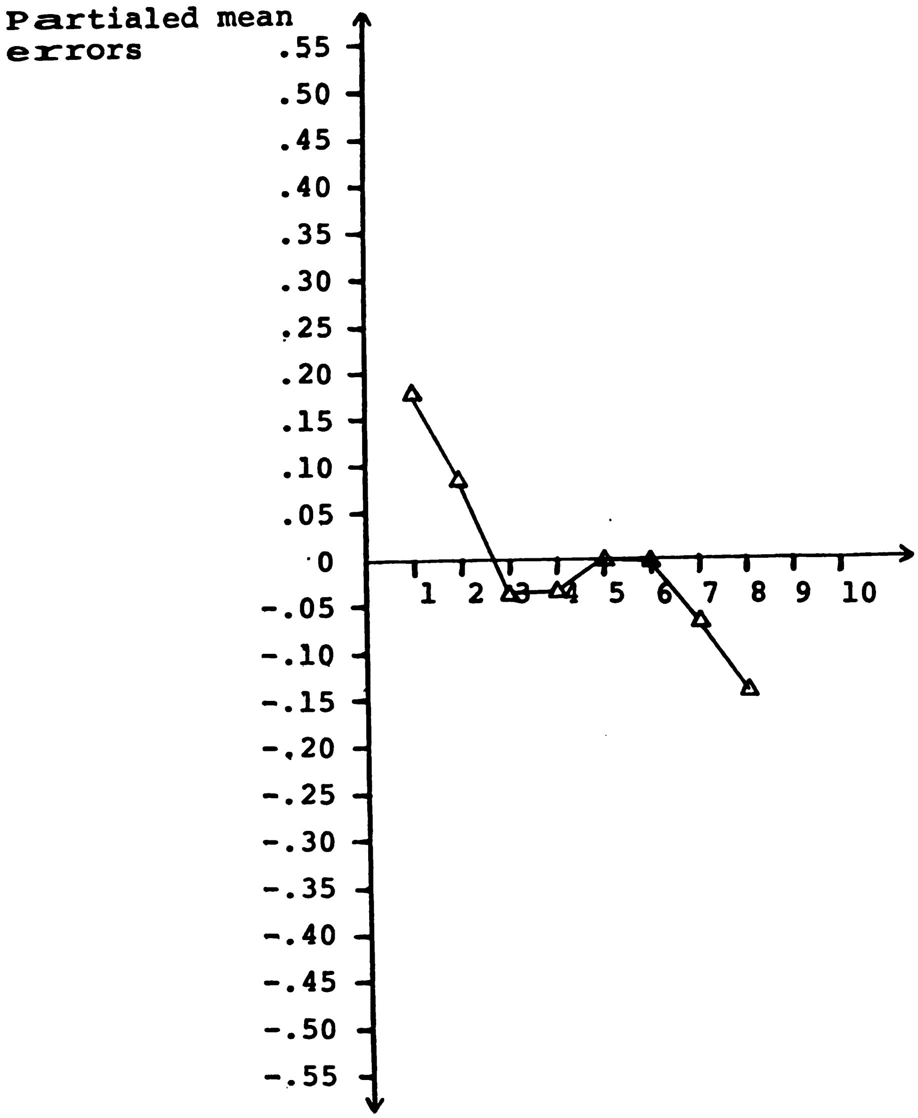


Figure 1

Subtest III: Partialed Mean Errors as a Function of Sequence

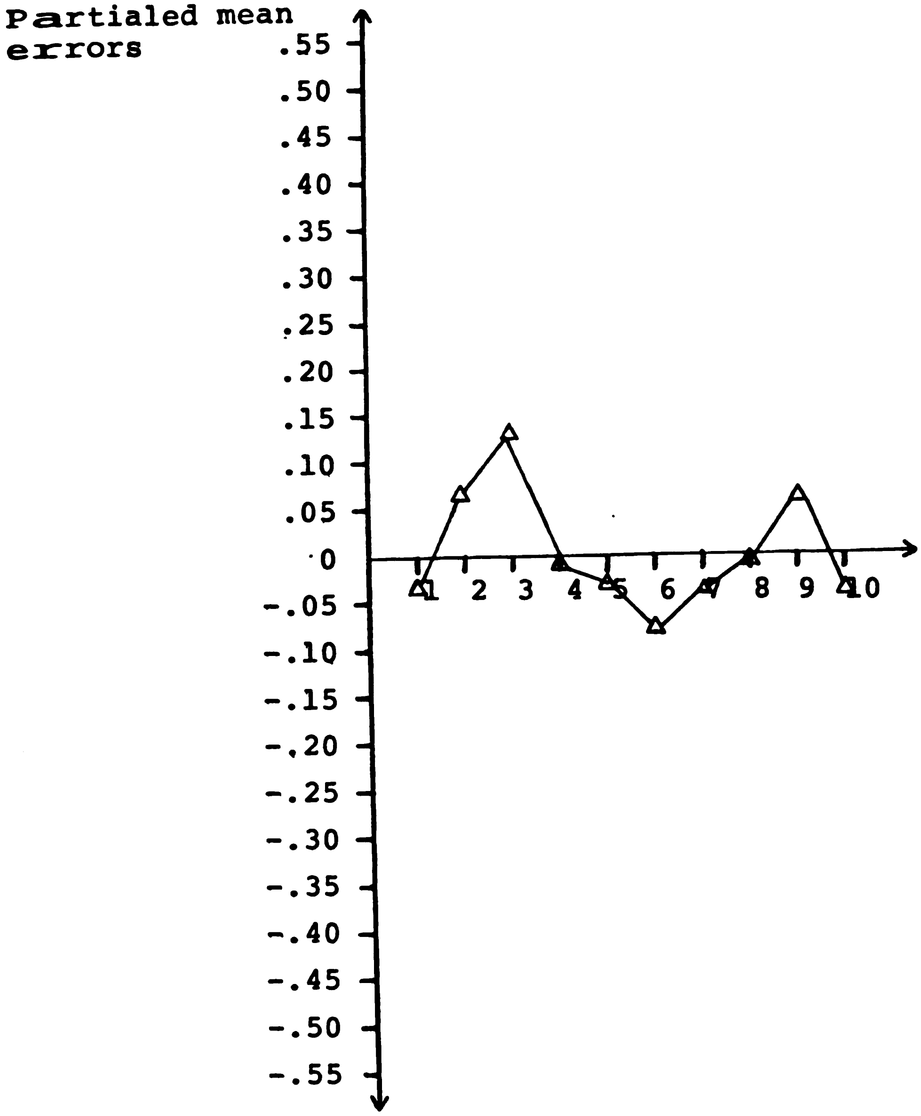


Figure 2

Subtest IV: Partialed Mean Errors as a Function of Sequence

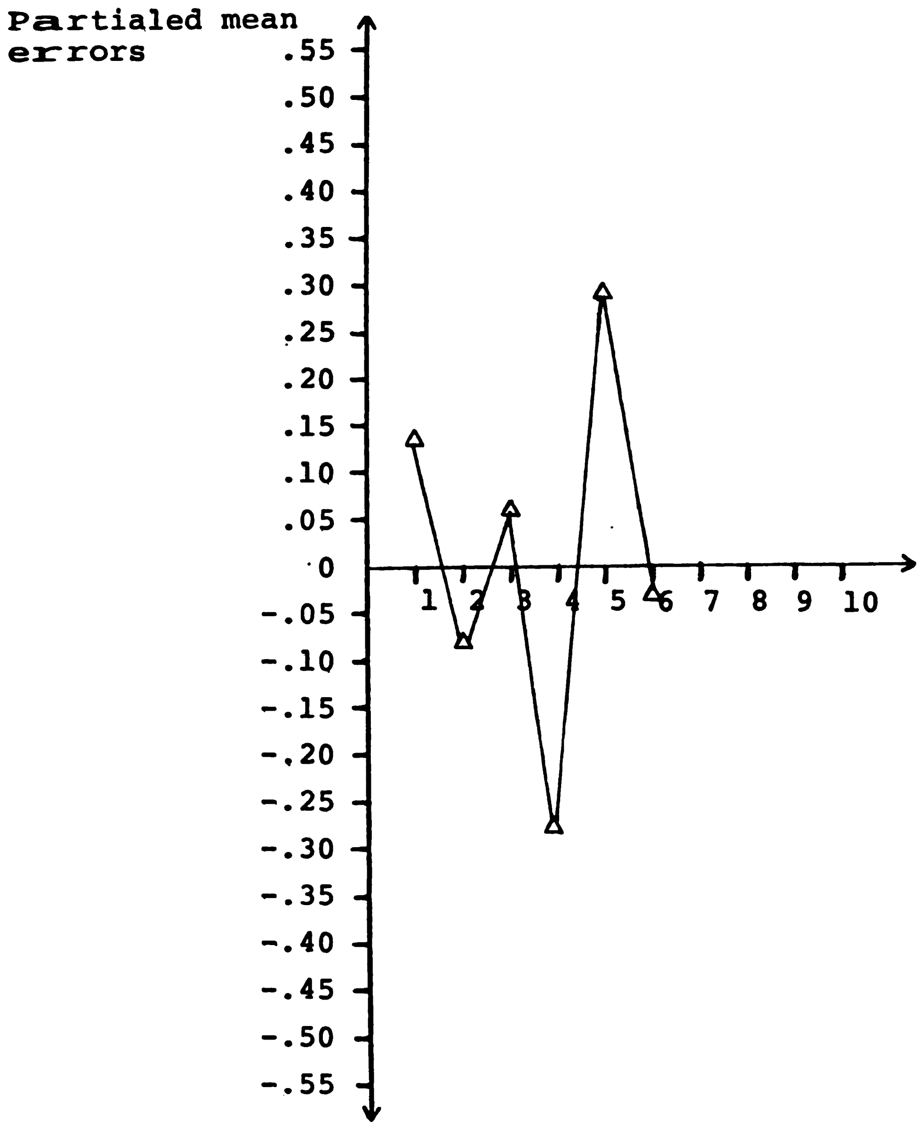


Figure 3

Subtest V: Partialed Mean Errors as a Function of Sequence

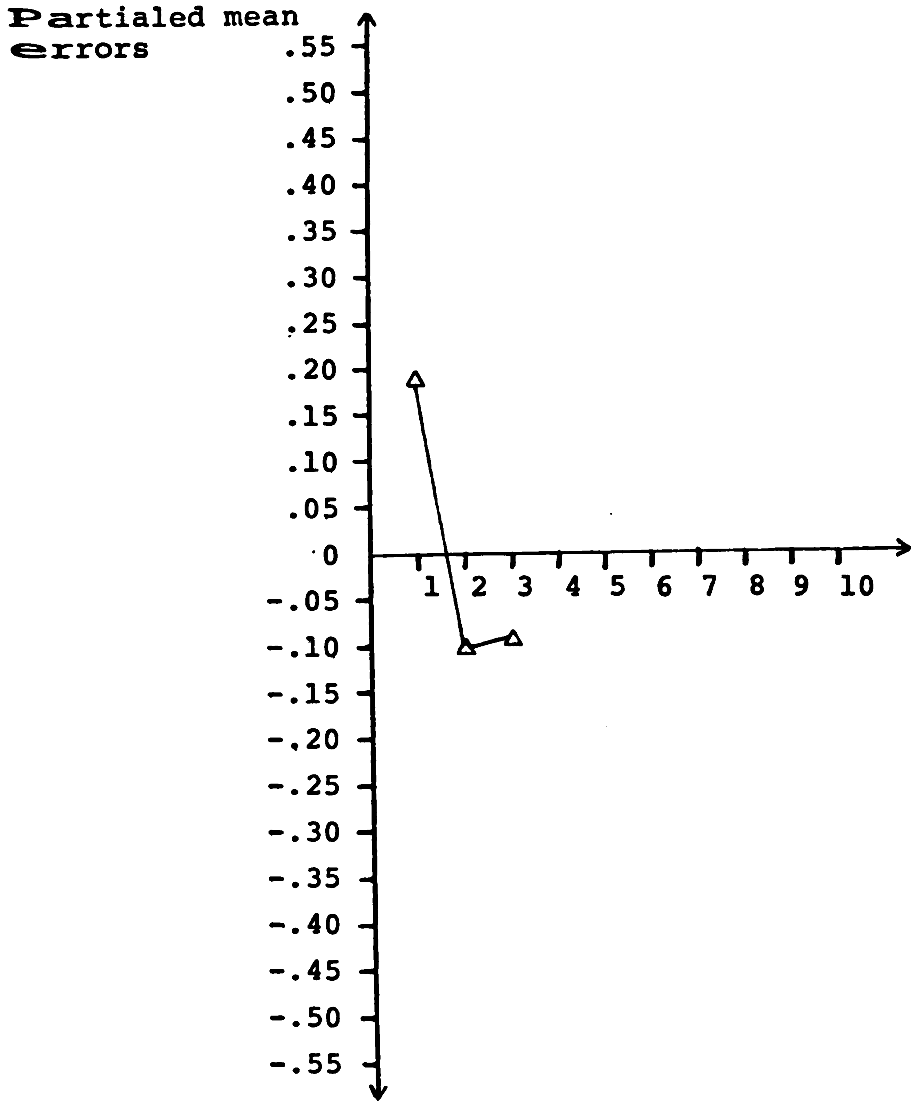


Figure 4

Subtest VI: Partialed Mean Errors as a Function of Sequence principle. Previously, only a single stimulus distinction had existed. The introduction of this distractor obviously rendered items temporarily more difficult, and after the eight trials represented by sequences 5 and 6, then the predicted negative slope resumed, across sequences 7 and 8.

The perturbation apparent between sequences 4 and 7 gave the curve a form highly compatible with the function, $y = x^3$ and for this reason, the cubic component for sequence emerged as significant. For more or less the same reason, albeit to a minimal extent, the quadratic component for sequence also proved significant. That is to say, the relatively asymptotic character of the curve at sequences 3 and 4 essentially introduced a quadratic component, but because of the temporary upward turn in the curve, the cubic approximation demonstrated a superior fit. In spite of the relevance and significance of the quadratic and cubic components, however, the linear approximation clearly manifested the better fit to the actual function.

Sequence accounted for but 1.9% of within-subjects variance in subtest IV, and though this was significant, it was not especially dramatic. Examinations of Figure 2 revealed that subjects found the items of sequence 1 easier than those of the next two sequences, which were experienced as progressively more difficult. From Figure I-D, it was inferred that the first six items (sequence 1) held far more information from which the correct organizing principle could be derived than did the following seven items (sequences 2 and 3), which consequently increased their difficulty. The succeeding 11 items comprising sequences 4, 5, and 6 were quite similar to those of sequences 2 and 3, and the negatively sloping curve in Figure 2 suggests that subjects perceived in this way, and gradually mastered the organizing principle. Then, the item design shifted again somewhat for sequence 7, and again for sequence 8, and both shifts affected the items by making them more difficult. Subjects, once more, apparently adapted to these changes, and items were experienced as somewhat easier in sequences 9 and 10.

The shifts in slope from positive to negative again permitted the introduction and significance of the cubic, though not the quadratic component, for sequence, although its importance was less dramatic than in subtest III. Because of the irregularity of the function as a whole, it was not possible to infer that the asymptotic characteristic of a learning curve had appeared at the locations on the graph at which its slope shifted. The significance of the linear component of sequence was in agreement with the observation that slope was generally negative, although only slightly so, and this supported the inference that learning, as predicted, had obtained, across sequences, although again, the finding was not so clear as its analog in subtest III.

Sequence accounted for 13.6% of the within-subjects variance in Subtest V, and this was far larger than subtests III and IV. The reason for this can be readily inferred from examining Figure 3: sequences differ radically in their average item difficulties. In a way analogous to subtest IV, though obviously remarkably more pronounced, the shift in item means across sequences was not as predicted. Indeed, the moderately significant linear component, accounting for .5% of within-subjects variance, actually manifested a positive slope, in contradiction to what was predicted.

To the quadratic component was attributed 1.87% of the within-subjects variance, and again, this in no simple way demonstrated learning, but rather only emphasized the unusual irregularity of the curve. In this case, too, the slope of the curve was clearly in the direction opposite to that predicted. In a similar way, the cubic component attained significance, accounting for a minute .3% of within-subjects variance. Had higher order orthogonal polynomials have been entered, it is highly likely that they should have captured sufficient variance to attain significance, as well.

By appealing to Figure I-E, in the Appendix, an interpretation of the rather complex curve in Figure 3 was made possible. The negative slope between sequences 1 and 2 was attributed to the structural similarity of their item stimuli, and this was in support of the hypothesis that learning would occur as a function of familiarity with items within the subtests. Sequence 3 involved a change in item design, and this was associated, in a way by now quite predictable, with an increase in item difficulty. Sequence 4 again proved less difficult, and this was by far the simplest of the sequences. The reason for this was readily derived from scanning items 26 through 33 in subtest V, and noting how transparent the organizing principle was in these items. All that was required of the subject was that she or he count the number of solid line segments. Sequence 5 was attended by a sudden and remarkable increase in errors, and this was attributed to the complexity of items 34 through 37, relative to items 26 through 33. Counting solid line segments in sequence 5 did not readily produce a correct response. Finally, the three items in sequence 6 again permitted the success of a rather straightforward counting strategy, and this was associated with a clear decrease in errors.

In summary of subtest V, while examination of the structural features of the item stimuli permitted lucid interpretation of the sequence-dependent, remarkable alternations in item difficulty, only in the case of sequences 1 and 2 can learning, as hypothesized, be argued to have convincingly occurred.

In subtest VI, again, sequence accounted for rather a dramatic proportion of the within-subjects variance, in this case, 13.7%. From scrutinizing Figure 4 and the polynomial components in Table 16, it was readily concluded that both linear (10.0%) and quadratic (3.7%) components were salient and significant. Had it been possible to include a cubic component, it might well have accounted for additional variance. The slope of the function in Figure 4 was negative, or in the predicted direction, and an examination of Figure I-F, in the Appendix, supported the notions that even substantial alterations in the structural aspects of the item stimuli failed to increase item difficulty. It was inferred that learning had indeed obtained.

<u>Trial</u>

This effect was also strikingly significant for each of the subtests, although it accounted for only about one half to one tenth the amount of variance attributed to sequence. The more moderate effect sizes were associated with decreased estimates of power. Table 19 includes the marginal means for trial, for each of the subtests, and computed with either raw item data, or following the partialing away of both between-subjects and option-attributable variance.

As was noted with sequence, the predicted, monotonic decrease in HCT item errors as a function of Trial, did not unequivocally obtain. Learning did indeed occur, and this conclusion was carefully justified, but the phenomenon was more complex than anticipated, because of the characteristics built into the HCT, and out, by implication, of the control of the present investigator. The partialed means were graphed as a function of Trial, within subtests, and the resulting curves may be viewed in Figures 5 through 8.

On the basis of Table 19 and Figure 5, the likelihood of errors, as a function of trial, was concluded to have shifted in the direction predicted. Thus, the slope of the curve obtaining was negative, and it began to assume an

Table 19. Marginal means for trial.

Derived With Raw Item Data

Subtest	Trial								
	1	2	3	4	5	6	7	8	9
III	.604	.515	.489	.533	--	--	--	--	--
IV	.469	.417	.484	.363	.329	.361	--	--	--
V	.498	.508	.468	.462	.370	.239	.411	.336	.436
VI	.300	.180	.253	.169	.229	.219	.260	.174	--

Derived With Partialed Data

Subtest	Trial								
	1	2	3	4	5	6	7	8	9
III	.076	-.023	-.054	.000	--	--	--	--	--
IV	.033	-.001	.017	-.035	-.061	-.150	--	--	--
V	.060	.087	.040	.036	-.055	-.153	-.045	.084	.043
VI	.085	-.002	-.015	-.012	.007	-.048	.047	-.061	--

Note. Variance attributable both to between-subjects differences and to correct option was partialed from raw item responses in preparing quantities in the second half of this table. Resulting values are standard scores rather than prohabilities of erroneous responses.

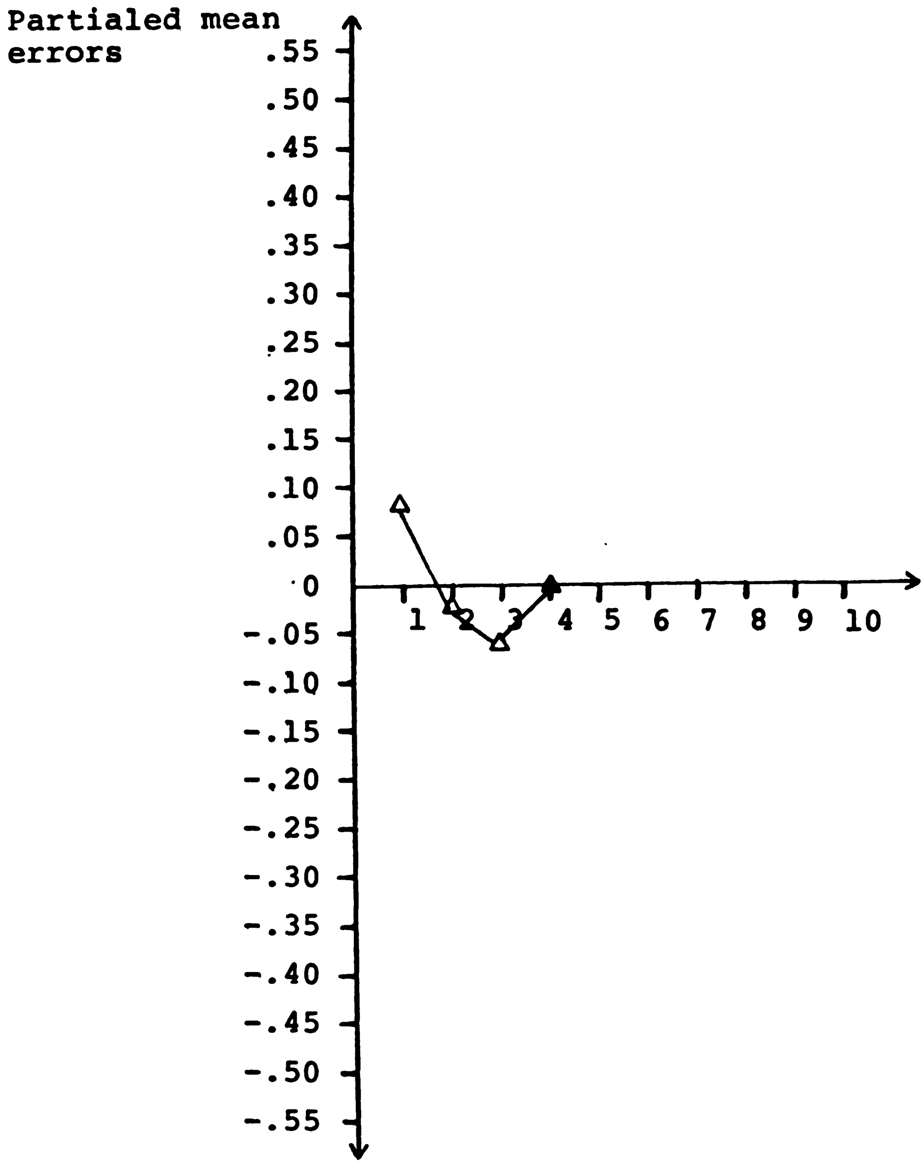


Figure 5

Subtest III: Partialed Mean Errors as a Function of Trial

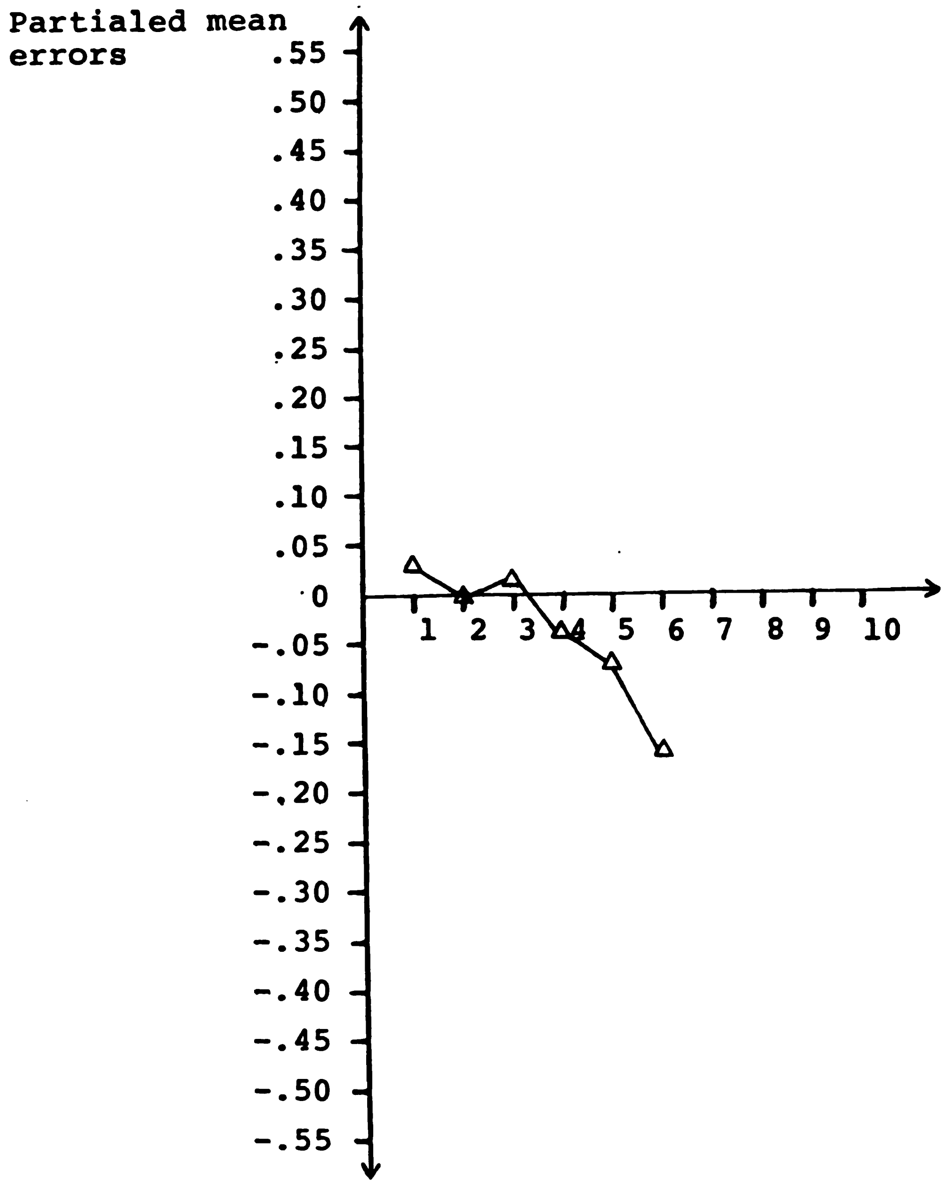


Figure 6

Subtest IV: Partialed Mean Errors as a Function of Trial

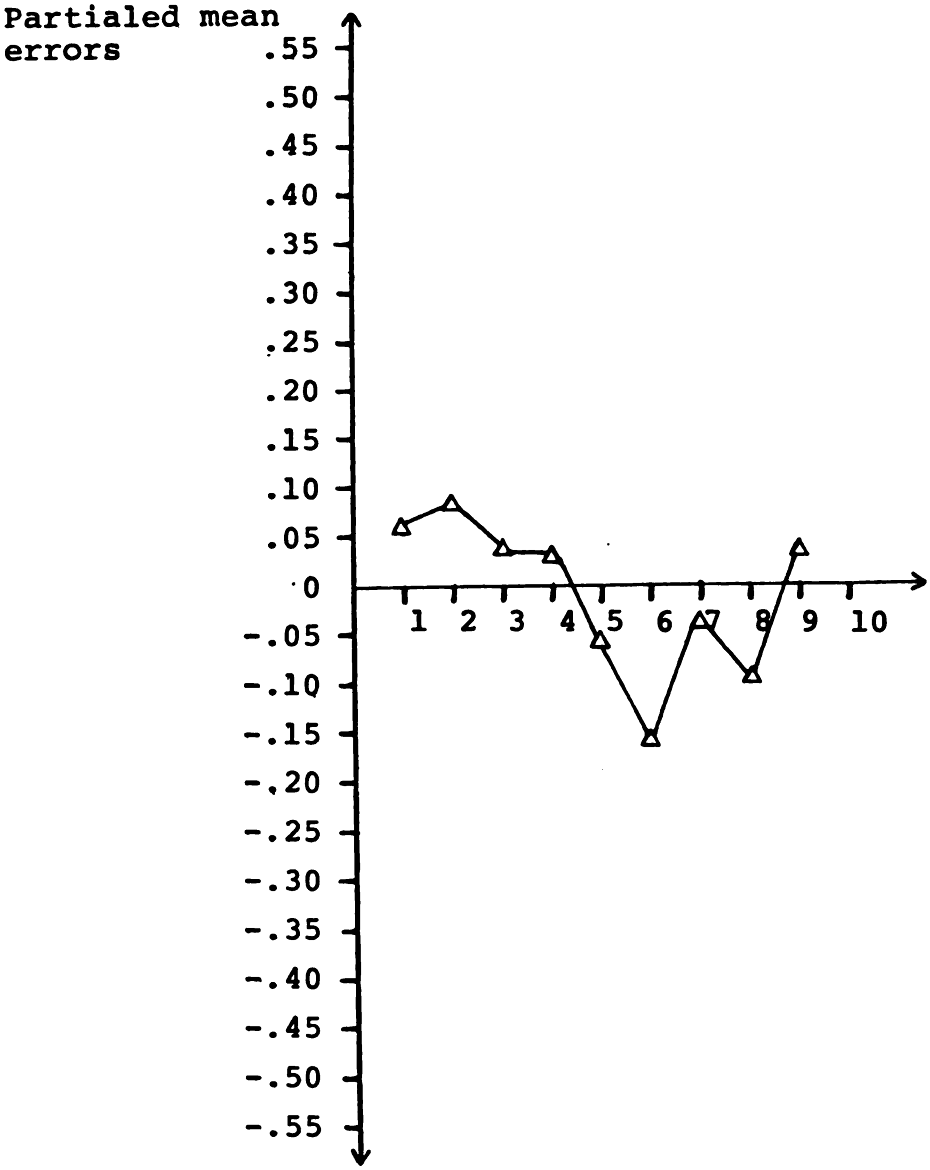


Figure 7

Subtest V: Partialed Mean Errors as a Function of Trial

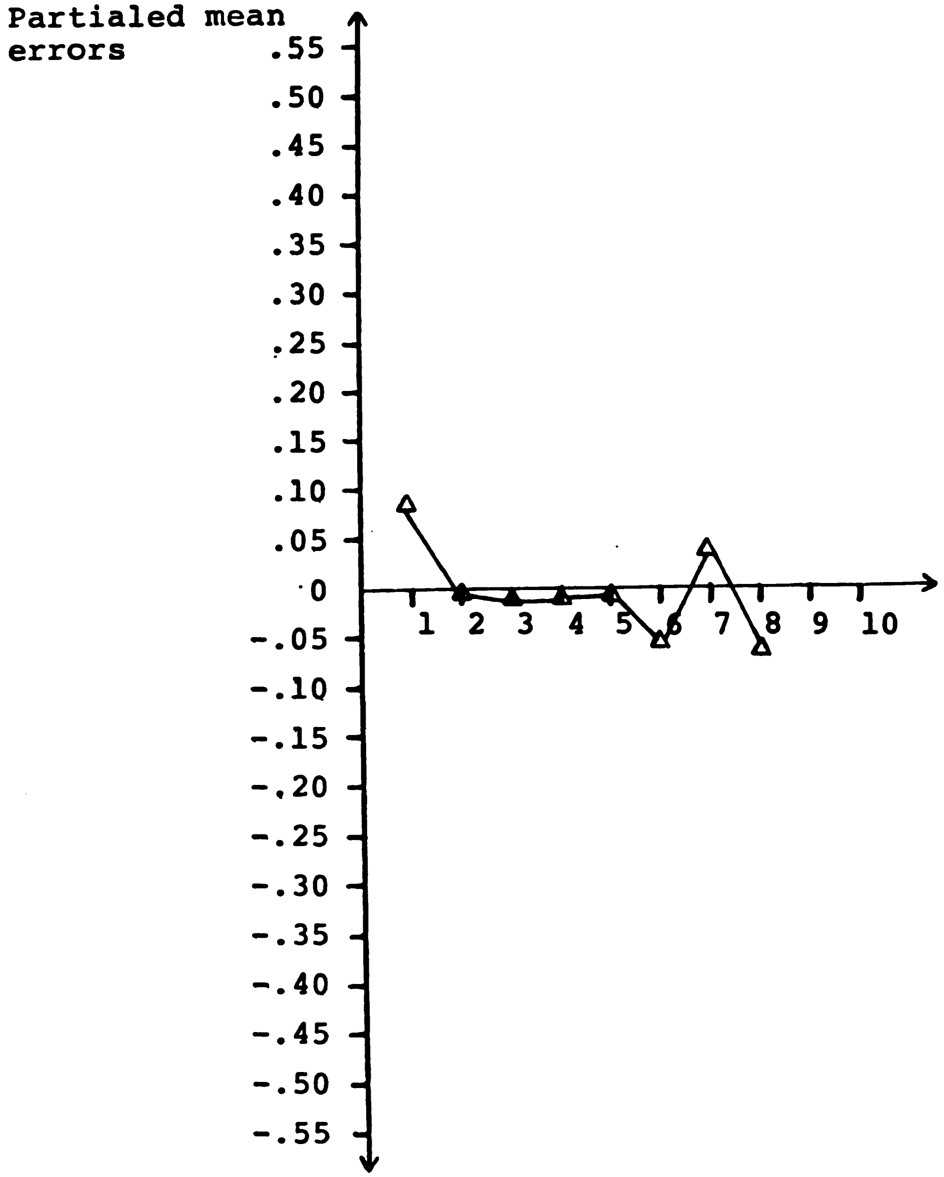


Figure 8

Subtest VI: Partialed Mean Errors as a Function of Trial asymptotic form between trials 2 and 3. The upturning of the curve between trials 3 and 4 was not, however, expected, and amounts to an increase in errors for trial 4 relative to trial 3 (or trial 2, for that matter). The reason for this upturning was very likely that the distributions of options across trials was nonrandom (see remarks concerning the interaction between option and trial, below), with trial 4 being loaded, so to speak, more heavily with options 2 and 4 than were the other trials. Because these options were more difficult than the other ones, trial four items were consequently rendered more likely to elicit errors than the others.

The leveling and upturning character of the curve was consistent with a quadratic function, and for this reason both linear and second degree orthogonal polynomial components proved significant. The quadratic component, however, augmented the more substantial (3.4%) linear element by a scant .9%. The cubic component added nothing whatsoever to the prediction.

As is depicted very nicely in Figure 6, the curve relating partialed errors to Trial in subtest IV was, as predicted, of negative slope. For the basic within-subjects variance or regression analysis, trial accounted for .9% of the available variance, and as indicated by Table 12, the curve was very well approximated by a linear component, with slight cubic curvilinearity introduced by the perturbation occurring between trials 2 and 3. The quadratic component proved irrelevant, and the linear and cubic elements combined accounted for about 90% of the variance attributable to trial.

The upward turn in the curve occurring at trial 3 was likely attributable to a preponderance, for this trial, of items with correct options 3 and 4, which were, for this subtest, the more difficult items. Nonetheless, it was clear that errors decreased monotonically as a function of Trial, much in the way hypothesized.

Subtest V produced an astonishingly complex curve, and one rather difficult to render amenable to present hypotheses. As can be deduced by examining Figure 7, the curve was relatively as expected for trials 1 through 6, but then suddenly the slope became positive for trials 7, 8, and 9. Also not as predicted, trial 1 was characterized by fewer errors than trial 2. Because of the irregular form of the curve, the cubic component emerged as significant. On the whole, the effect captured 1.7% of the within-subjects variance, and this was predomonantly linear in character with some improvement in fit accomplished by the addition of the cubic component.

The departure of trial 1 from expectation could not be accounted for by appealing to the distribution of correct options, for, by examining Figure 8-E in the Appendix, it was concluded that there were as many difficult items for this trial as easy ones, determined on the basis that is, of correct option. The partialed individual item means also were examined, but this, too, failed to clarify the finding.

The upward trend obtaining for trials 7, 8, and 9 was also rather difficult to interpret. Trial 7 was clearly heavily loaded with correct options 1 and 2, and was never defined by option 2, and consequently, it was not surprising that error frequency increased for this trial. Trial 8 manifested a decrease in errors relative to trial 7, but this was still a more difficult trial, on the average, than was predicted. The items comprising this marginal cell included two instances of correct option 3, and one each of correct options 1 and 2, and this composition did not seem to support its relative difficulty. Nor could further light be shed upon the matter by appealing to the structural characteristics of the relevant items, as they appear in Figure I-E. Finally, even an examination of Figure 3, with consideration being given to sequences (1, 2, 3, and 4) having a trial 8 failed to clarify the matter. That is, though these are the earlier sequences in the subtests, they are by no means the most difficult. Trial 9 was readily explained, as it was represented solely by sequence 1, that characterized by more errors than any other sequence, save number 6.

Subtest VI was somewhat more coherent than subtest V in the regard of trial, and this effect accounted for 1.7% of the within-subjects variance therein. Both linear and cubic components were important, as was also true for subtests IV and V. However, in this case, each component accounted for but 4% of the within-subjects variance, or approximately 24% of the variance allocated to trial, indicating that 52% of the available variation was attributable to aspects of trial not predicted.

From the curve appearing in Figure F, it can be seen that trial one was the more difficult of the lot, and that trials 2, 3, 4, and 5 were easier, and about equally difficult. Examination of the marginal means derived from raw item data revealed that not only was subtest VI by far the simplest of those analyzed, but also, trials 2, 7, and 4 were comprised of items with very low difficulty. It was believed possible that these trials illustrated the operation of an asymptotic process. Trial 6 manifested decreased difficulty, but this was attributed to the fact that of the three items composing it, two manifested correct option 3, and one, correct option 1, the easier two options for this subtest. Trial 7, on the other hand, included one item each of correct options 2 and 4, and consequently its frequency of errors increased. Trial 8, once again, was characterized by correct options 1 and 3. The decreases in errors for trials 6 and 8 were inferred to demonstrate further learning, relative to trials 1 through 5.

It was decided against employing the curvilinear components of sequence and trial in subsequent consideration of the two and three variable interactions. The reasons for this were that the linear components of these variables tended to account for far more variance than the curvilinear components, and that even when the curvilinear components proved relevant, it was generally because they permitted a better fit to curves distorted from linearity for reasons other than hypothesized. It was decided, consequently , to enter the dummy coded versions of trial and sequence into interactions.

Option by Sequence

Both the option by sequence and the option by trial interactions were predicted to reach significance, and the rationale behind each prediction was that the impact the item structures or characteristics of each subtest upon the types of errors made would shift in form over time, in a way consistent with learning. Insofar as error distributions across the various options was concerned, it was consequently anticipated that early in the subtests, some options would appear more difficult or more simple than others, and that these disparities would vanish as a function of time or familiarity with the subtests, as defined by the passage of sequences or trials. Convergence in apparent option difficulty levels was thus expected, over time.

Of the total set of interactions examined, option by sequence emerged the more potent, in that it readily attained significance for all four subtests examined. For subtest III, it accounted for 2.1% of within-subjects variance; for subtest IV, it accounted for 1.6%; for subtest V, 4.0%, and for subtest VI, 1.2%. Next to sequence itself, this effect tended to operate as a more important determinant of HCT item behavior than any other effect studied at the level of the within-subjects design. It had been predicted that this interaction would attain significance. It was argued that as familiarity with a subtest of the HCT increased, item characteristics would prove less distracting and hence less important in determining responses made. This shift, it was reasoned, would produce a diminution in the salience of correct option, as a predictor of HCT item behavior. Pertinent cell and marginal means have been reproduced in Table 20 for subtest III, Table 21 for subtest IV, Table 22 for subtest V, and Table 23 for subtest VI. As was the case previously, the strategy of partialing both between-subjects variance, and that variance attributable to option was applied here, as well. In order to facilitate the interpretation of these tabled valves, portialed mean errors were graphed as a function of sequence, for each of the subtests, and these curves were reproduced in Figures 9 through 12. Each figure contains four curves; one, that is, for each option.

The four curves depicted in Figure 9 all demonstrate the negative slope already noted in Figure 1. Variation very obviously occurred, but the majority of this was attributed to the lack of balance in the design, or more lucidly, to the nonrandom representation of trial at the various points on the curves. As well, various aspects of the HCT items themselves may well have exerted an uncontrolled impact upon mean errors. These sources of influence were ignored in interpreting the interaction.

Consonant with the shape of the curve in Figure 1, three of the four curves in Figure 9 bend upward, in the direction of greater mean errors, as the more complex

Table 20. Subtest III: Cell and Marginal Means for Option X Sequence

Sequence	Option				
	1	2	3	4	Marginal
1	.243	----	.092	.141	.179
2	-.034	.132	----	.109	.085
3	-.006	.068	-.091	----	-.030
4	----	-.022	-.055	-.016	-.027
5	-.086	----	.136	.036	.000
6	-.092	.061	----	-.051	-.006
7	----	-.195	.004	-.058	-.062
8	-.091	-.236	----	-.147	-.142
Marginal	.005	.012	-.076	.048	.000

Note. Entries have been partialed with respect to variance attributable to between-subjects differences and to option.

Table 21. Subtest IV: Cell and Marginal Means for Option X Sequence

Sequence	Option				
	1	2	3	4	Marginal
1	.087	-.061	-.099	-.108	-.033
2	-.068	.085	.161	--	.066
3	.077	--	.173	.128	.131
4	-.078	.046	-.032	.037	-.007
5	-.022	-.055	-.117	.076	-.029
6	-.144	-.049	--	-.017	-.070
7	--	-.026	-.053	.008	-.031
8	-.012	-.021	--	.001	-.011
9	.109	.002	.075	--	.065
10	--	-.008	-.131	-.010	-.040
Marginal	-.114	-.048	.085	.074	.000

Note. Entries have been partialed with respect to variance attributable to between-subjects differences or to option.

Table 22. Subtest V: Cell and Marginal Means for Option X Sequence

Sequence	Option				
	1	2	3	4	Marginal
1	.140	.017	.133	.154	.134
2	-.051	-.080	-.075	-.088	-.074
3	.112	-.062	-.062	.233	.056
4	-.282	-.218	-.238	-.352	-.272
5	.019	.522	.506	.124	.293
6	--	-.061	-.035	-.003	-.031
Marginal	.093	-.054	-.031	-.008	.000

Note. Entries are based upon means with variance due to between-subjects differences and option having been partialed.

Table 23. Subtest VI: Cell and Marginal Means for Option X Sequence

Sequence	Option				
	1	2	3	4	Marginal
1	.211	.122	.136	.357	.180
2	-.130	-.069	-.141	-.074	-.099
3	-.130	-.057	-.135	-.031	-.088
Marginal	.063	-.029	.035	-.063	.000

Note. Entries are means with variance due to between-subjects differences and option having been partialed.

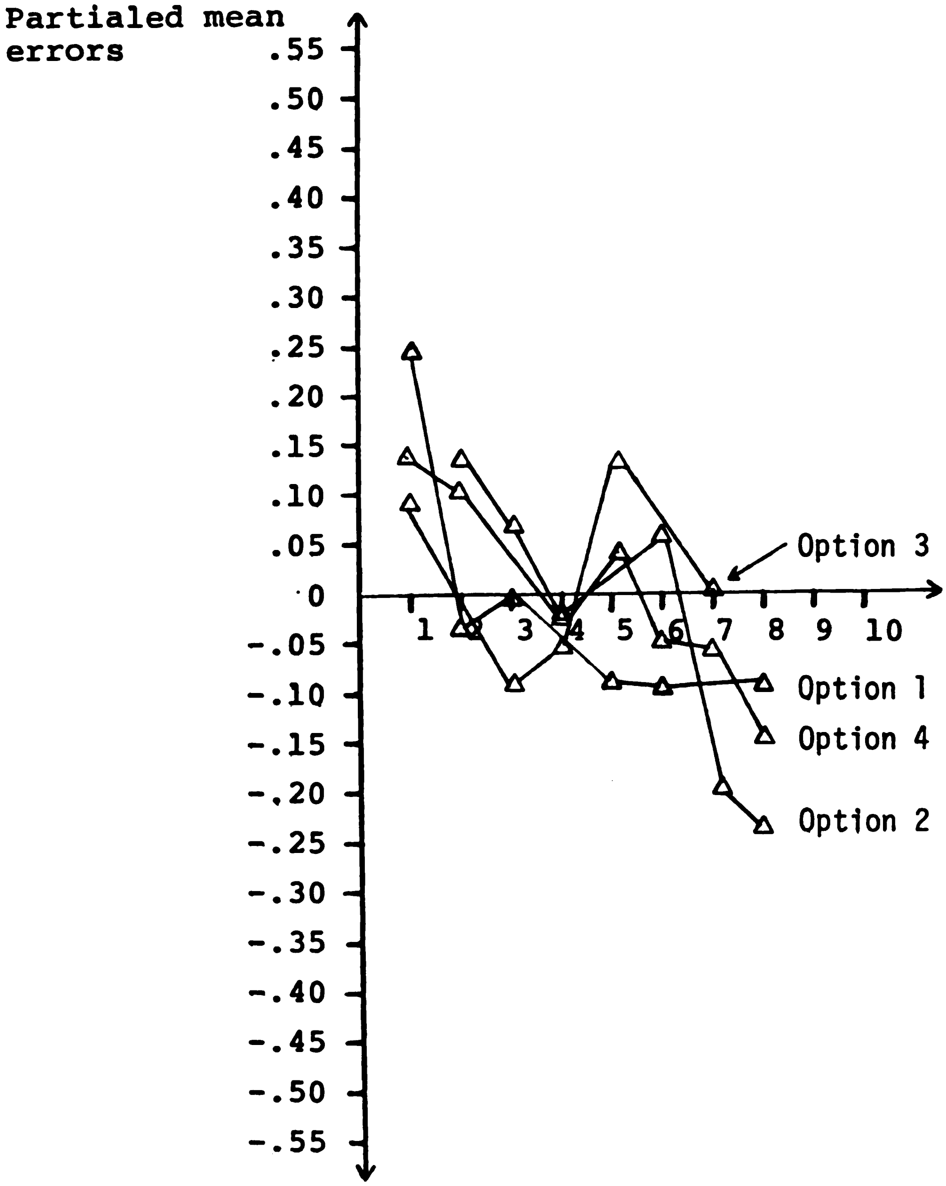


Figure 9

Subtest III: Option-Segregated Curves as a Function of Sequence

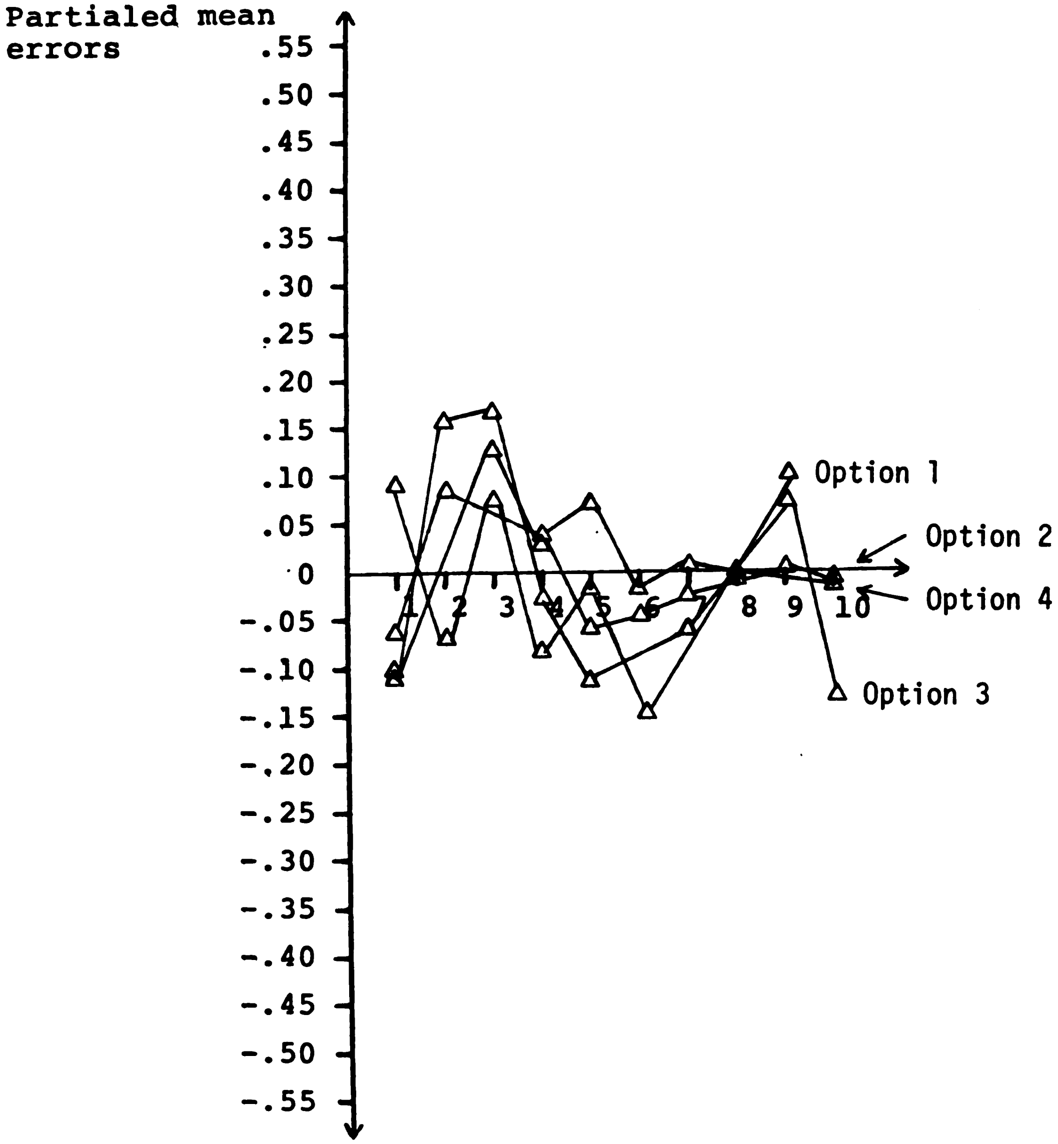


Figure 10

Subtest IV: Option-Segregated Curves as a Function of Sequence

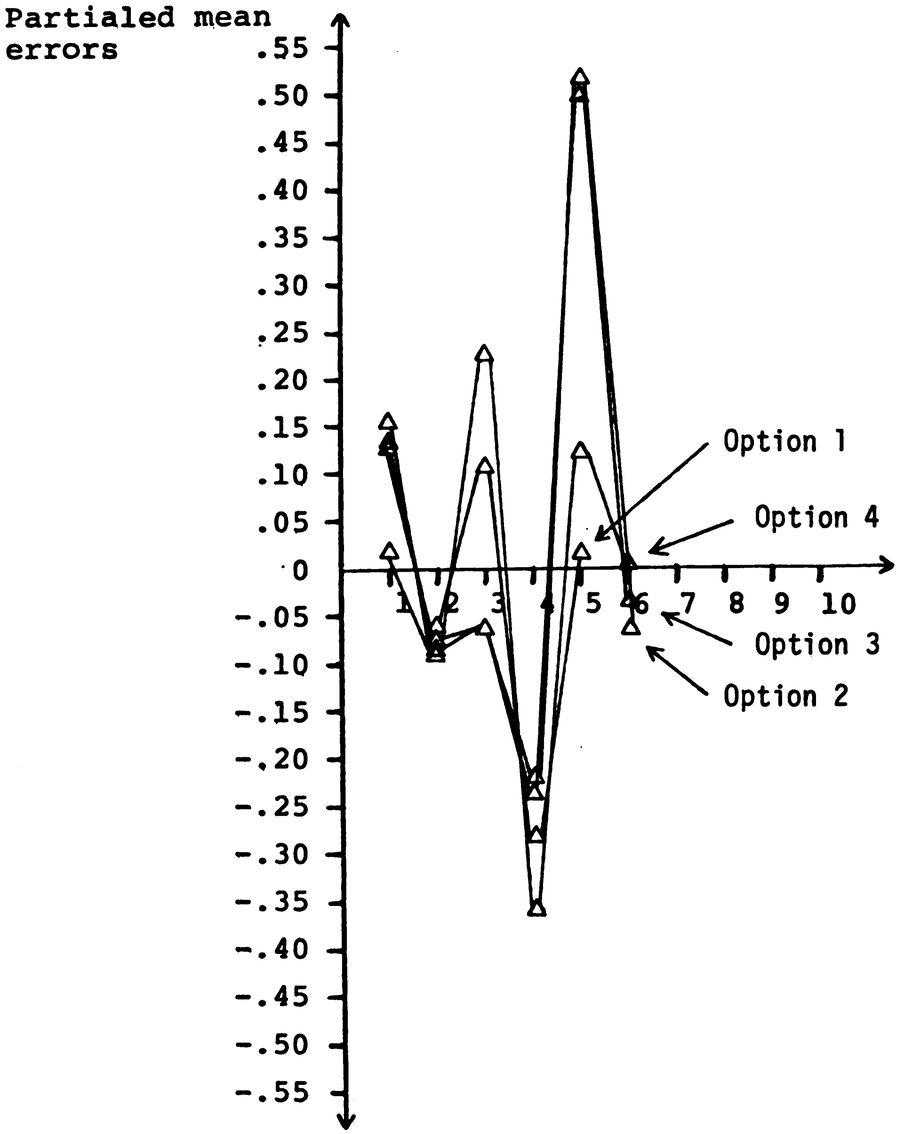


Figure 11

Subtest V: Option-Segregated Curves as a Function of Sequence

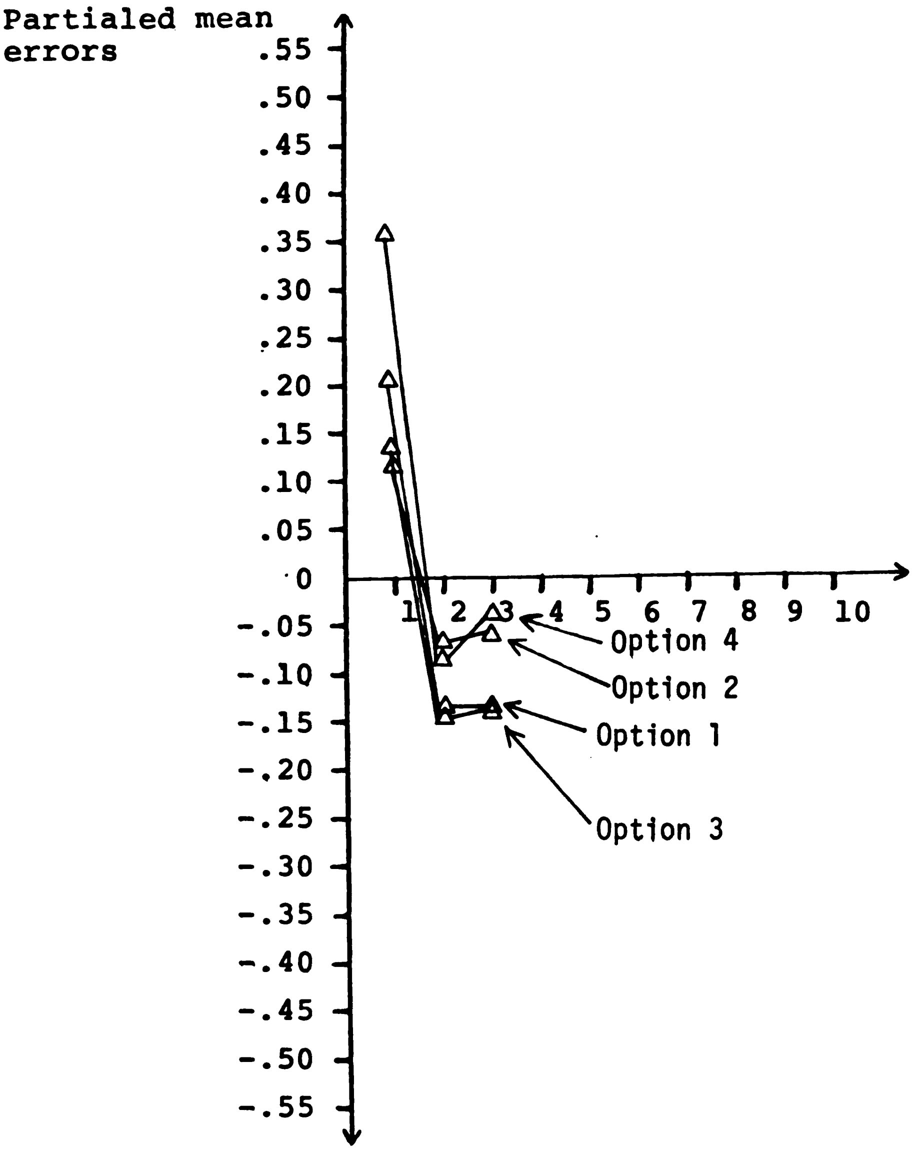


Figure 12

Subtest VI: Option-Segregated Curves as a Function of Sequence sequences 5 and 6 are encountered. The sole curve which failed to show this trend was that for option 1. A rather convincing explanation for this was arrived at upon examining Figure I-C, in the Appendix. Items 17 through 20 form sequence 5. It will be recalled that options 1 and 3 were favored by subjects for this subtest, because stimuli tended to be divisible into two groups, one always containing three events, and one, a single event. Then, by counting, subjects tended to arrive at a response of either "one" or "three". This produced a bias in the direction of fewer errors for these options, although correct responses were made for inaccurate reasons. The negative slopes noted for the curves in Figure 9 across sequences 1 through 4 indicate that this uninsightful response set was relinquished by subjects as they were punished. Then, when sequence 5 was encountered, and it again became difficult to divine the accurate basis upon which to respond, subjects were very likely pushed, as it were, in the direction of counting once again. In this case, however, the first three items of the sequence offered only punishment as a consequence to choosing option 3. Option 1, on the other hand, was liberally reinforced, as it appears twice as the correct solution during the first three items of sequence 5. By the time correct option 3 arrived, with item 20, this response had been extinguished, and many subjects erroneously selected some other option, including, quite possibly, option 1. Sequence 6 contained no instance of correct option 3, and the rather high error rate noted for this option in sequence

7 probably indicates that subjects never recovered their faith in this response.

The combination of lack of trust in option 3, and the increased complexity of the items beyond sequence 4, also seemed to have affected the error rates for options 2 and 4. On the one hand, that is, the punishment of option 3 encouraged subjects to try, instead, options 2 or 4, even though in the earlier sequences these alternatives could not be arrived at solely by utilizing the strategy of item counting. However, the more complex, later sequences also are characterized by stimuli which differ more from one another, and consequently this may well have encouraged subjects to arrive at a "four" response by counting. Finally, sequences 7 and 8 contain items which can quite readily be separated into two groups of stimuli, and in all likelihood this accounted for the great reduction of errors at sequence 8 for the option, 2.

In summarizing the option by sequence interaction for subtest III, it was concluded that its significance did not indicate that familiarity with the subtest had shifted the response bias introduced by item characteristics in the direction of more insightful behavior. Rather, this interaction revealed that aside from the general decrease in errors as a function of familiarity with the subtest, the shifts in difficulty, over time, of various options, was attributable to alterations in the content or structure of items. In particular, as the accurate organizing principle again became obscure, counting was resorted to. Moreover, as alternate, though equally incorrect counting strategies became available, they were used.

In comparing Figures 2 and 10, it can be deduced that to a fair approximation, with exceptions as noted below, the family of option-segregated curves for subtest IV behaved in a fairly coherent way. On the whole, the slopes of these curves were negative, indicating a general decrease in errors, as a function of familiarity with the subtest, and largely irrespective of correct option.

The first important exception was that for sequence 1, option 1 proved quite difficult relative to the remaining options, which were roughly equally difficult. For sequence 2 the pattern shifted, with option 1 manifesting fewer mean errors than the others. The higher frequency of errors for correct option 1 items in sequence 1 was inferred, after Simmel and Counts (1957, pp. 51-71), to have come about as a consequence of the subjects' difficulty to arrive at a "one" response by counting some aspect of the stimuli in sequence 1. As much was deduced by examining Figure I-D. Shunning option 1, subjects made many errors when this alternative was actually the correct one. The relative decrease in errors noted for items in sequence 1 with correct options 3 and 4 was attributed to the "success" of counting strategies, albeit for inaccurate reasons. That option 2 fared so well in its error rate was probably due to its occuring rather late in the sequence, at the position of trial 5, by which time many subjects had divined the correct organizing principle.

Upon the arrival of sequence 2, the stimulus array shifted such that the cuing numbers apparent in sequence 1 items were no longer present. Errors for option 1 decreased significantly, but errors for the other options increased just as dramatically. Based on the appearance of the item stimuli, these changes were inferred to have come about because suddenly the stimuli were quite appropriately viewed as unitary constructs, calling for a "one" response. Beyond this point, on the basis of the curves in Figure 10, it would appear that items with correct options 3 and 4 tended to remain more difficult than items with correct options 1 and 2. As well, whenever the stimulus figures were closed, or manifested an unbroken line completely enclosing an inner space, then items with correct option 1 manifested decreased mean errors. The reason for the first of these trends was that, as was mentioned above while discussing the effect, 'Option', quadrants 3 and 4, associated with correct options 3 and 4, were counterintuitively placed with respect to one another and to quadrants 1 and 2. Thus, subjects invariably tended to confuse these quadrants, and the associated mean error values remained inflated. The second trend was explained as the predisposition by subjects to emit a "one" response when the stimulus figure could be viewed as a single, coherent event. This proved possible for sequences 2 (items 8 through 10), 4 (items 14 through 17), 6 (items 22 through 24), and to a lesser extent, 8 (items 29 through 32). For sequences which did not facilitate the perception of stimuli as single, coherent objects, items with correct option 1 tended to manifest somewhat inflated mean error scores, and this was concluded to indicate that subjects had abandoned the "one" response, with the consequence that items for which it was the correct choice manifested elevated errors.

As was already noted in discussing this interaction for subtest III, option by sequence can be said to have demonstrated appreciable learning effects in the case of subtest IV. Rather, the interplay of option and sequence was concluded to have arisen as a consequence of alterations in item characteristics which tended to enhance or suppress the likelihood of response options, and in turn, which irrelevantly, insofar as learning was concerned, inflated or deflated mean error scores.

For subtest V, the array of findings was somewhat less complex to interpret for the option by sequence interaction than was true of subtests III and IV. In general, the curves in Figure 11 very neatly followed their option-unsegregated analog in Figure 3. The sole remarkable discrepancy was that items with correct options 1 and 4 tended to produce nearly identical mean error scores across sequences, and the same was true of items with correct options 2 and 3. The curves for these two distinct pairs of items converged, more or less, at sequences 1, 2, 4, and 6, and were sharply divergent at sequences 3 and 5. At sequence 1, which was otherwise strongly convergent across options, items with correct option 2 manifested a lower mean error score than items with one of the other three options correct. In all likelihood, this was because those items with correct option 2 could not be responded to on the basis of the set which had been acquired during the previous subtest, number IV. From scanning items 1 through 9 in Figure I-E, it was determined that those items with option 2 as correct could not be solved by appealing to the "quadrant" schema as learned through contact with the preceding subtest. Items 4 and 6 also did not quite fit the previously learned organizing principle, but these probably strongly elicited "one" responses, as they so clearly were unitary, coherent stimuli. Items characterized by correct option 2 were most likely to be correctly solved, then, because they did not so readily elicit an erroneous principle from subjects.

The divergent locales on the curves obtaining at sequences 3 and 5 were explained, again, by appealing to the stimuli as depicted in Figure I-E. For sequence 3, the items with correct option 4 in all likelihood tended to be seen as a pair of line segments, and these elicited "two" responses (Simmel and Counts, 1957, pp. 72-96), for this reason. Those items with correct option 1 tended to elicit either "two" or "three" responses. The explanation for the choice of option 2 was obvious. It proved more difficult to understand why in a situation of ambiguity a choice of "three" would prevail over a choice of "one", although this was also noted to have occurred in response to the majority of the items in subtest III. Sequence 5 produced the highest mean frequency of errors of any group of items in this subtest or any other. In fact, items 35 and 36 of this subtest were the most difficult items analyzed from this or any other subtest. The reason for this was in all likelihood an overwhelming inclination experienced by subjects to view the stimuli in sequence 5 as whole objects, and to emit "one" responses. The response of "four" was also apparently encouraged, due to its reduced mean error score, but this item was the last one in the sequence, and by that time, many subjects had probably solved the special problem posed by sequence 5. A careful examination of items 35 and 36 also suggested that in addition to "one" responses, these items were probably encouraging, respectively, of "two" and "four" responses. With sequence 6, error frequencies again dropped for correct options 2 and 3.

The interaction involving option and sequence, for subtest V, was again concluded not to have supported the hypothesis that learning would operate to decrease the impact of item characteristics upon subjects' behavior. Instead, the interaction in this case was quite explicitly a function of the impact changes in item structure had upon the likelihood of one option being selected over another. This could not be said to have anything whatsoever to do with learning.

The family of curves depicted in Figure 12 follow their generic analog in Figure 4 very closely. There were few divergent aspects to the curves, and when present, these were readily attributed to item characteristics. Thus, option 1 items proved more difficult for sequence 1 than did option 2 items, and this pattern shifted for sequences 2 and 3. The early (i.e., sequence 1) juxtaposition of mean error scores for options 1 and 2 was probably attributable to a "resurfacing," as it were, of the organizing principle learned from subtest IV. Then, during sequence 2, item 16 probably was generative of erroneous "two" responses, while the items manifesting correct option 1 were probably less likely to facilitate errors. For sequence 3, it was difficult to understand or offer an account concerning why items defined by correct options 2 and 4 proved more difficult, on the average, than those manifesting correct options 1 and 3. Rather, it would have seemed more plausible that items with correct options 1 and 3 would have been readily confused with one another, producing higher mean error scores. The sole reason the current investigator was able to arrive at was that items early in this sequence tended to be characterized by correct options 2 and 4, and consequently, perhaps more mistakes were made with these, before the sequence had been mastered.

For all four of the subtests, it was concluded that although the option by sequence interactions were salient and significant determinants of HCT item behavior, in no circumstance could it be argued that the nature of this influence was in the direction of item characteristics proving less distracting or disruptive over time (i.e., sequences). Thus, it was inferred that learning could not be demonstrated to have manifested a moderating influence upon the relationships between item characteristics and mean error scores. On the other hand, the relationships between item characteristics, overall error rates, and option-dependent error rates, were elucidated significantly by consideration of the option by sequence interactions, and this further emphasized the salience of item characteristics, and particularly item complexity, upon HCT item errors.

Option by Trial

For three of the four HCT subtests considered, the option by trial interaction also emerged as significant. Thus, the interaction accounted for .2% of the within-subjects variance for subtest III, .7% for subtest V, and .3% for subtest VI. As was true of the main effects involving sequence, relative to trial, these interactions captured dramatically less variance than did their counterparts involving sequence. Cell and marginal partialed mean error scores for the three subtests manifesting significance appear in Tables 24, 25, and 26. As was the practice with the option by sequence interaction, option-segregated families of curves were drawn, and these are included in Figures 13, 14, and 15. Results were not documented in detail for subtest IV, as the option x trial interaction failed, for this subtest, to reach significance, and an examination of the partialed mean error scores proved not to provide more information concerning the operation of option and trial than had already been made available by considering the main effects for these variables.

Table 24. Subtest III: Cell and Marginal Means for Option X Trial

Sequence	Option 1	2	3	4	Marginal
1	.088	.183	-.003	-.013	.076
2	-.092	-.064	.019	.000	-.023
3	-.002	-.129	-.087	-.039	-.054
4	-.083	.010	.027	.045	.000
Marginal	.055	.012	-.076	.048	.000

Note. Entries were partialed with regard to between-subjects and option variance.

Table 25. Subtest V: Cell and Marginal Means for Option X Trial

Sequence	Option				
	1	2	3	4	Marginal
1	.090	-.094	-.035	.405	.060
2	.172	--	.221	-.161	.087
3	-.010	.231	-.103	--	.040
4	-.281	.029	--	.144	.036
5	.053	.180	-.085	-.372	-.055
6	--	-.220	--	-.087	-.153
7	.159	-.122	--	.063	-.045
8	.019	.034	-.196	--	-.084
9	--	--	.043	--	.043
Marginal	.093	-.054	-.031	.008	.000

Note. Entries are mean error scores with variance attributable to between-subjects differences or to option having been partialed.

Table 26. Subtest VI: Cell and Marginal Means for Option X Trial

Sequence	Option 1	2	3	4	Marginal
1	.302	.022	--	-.071	.085
2	--	.014	--	-.032	-.002
3	-.095	--	-.025	--	-.015
4	--	.019	--	-.077	-.012
5	-.206	--	.160	-.086	.007
6	.120	--	-.132	--	-.048
7	.130	-.086	--	.357	.047
8	-.143	--	.053	-.093	-.061
Marginal	.063	-.029	.035	-.063	.000

Note. Entries are partialed mean error scores, with regard to between-subjects and option-attributable variance.

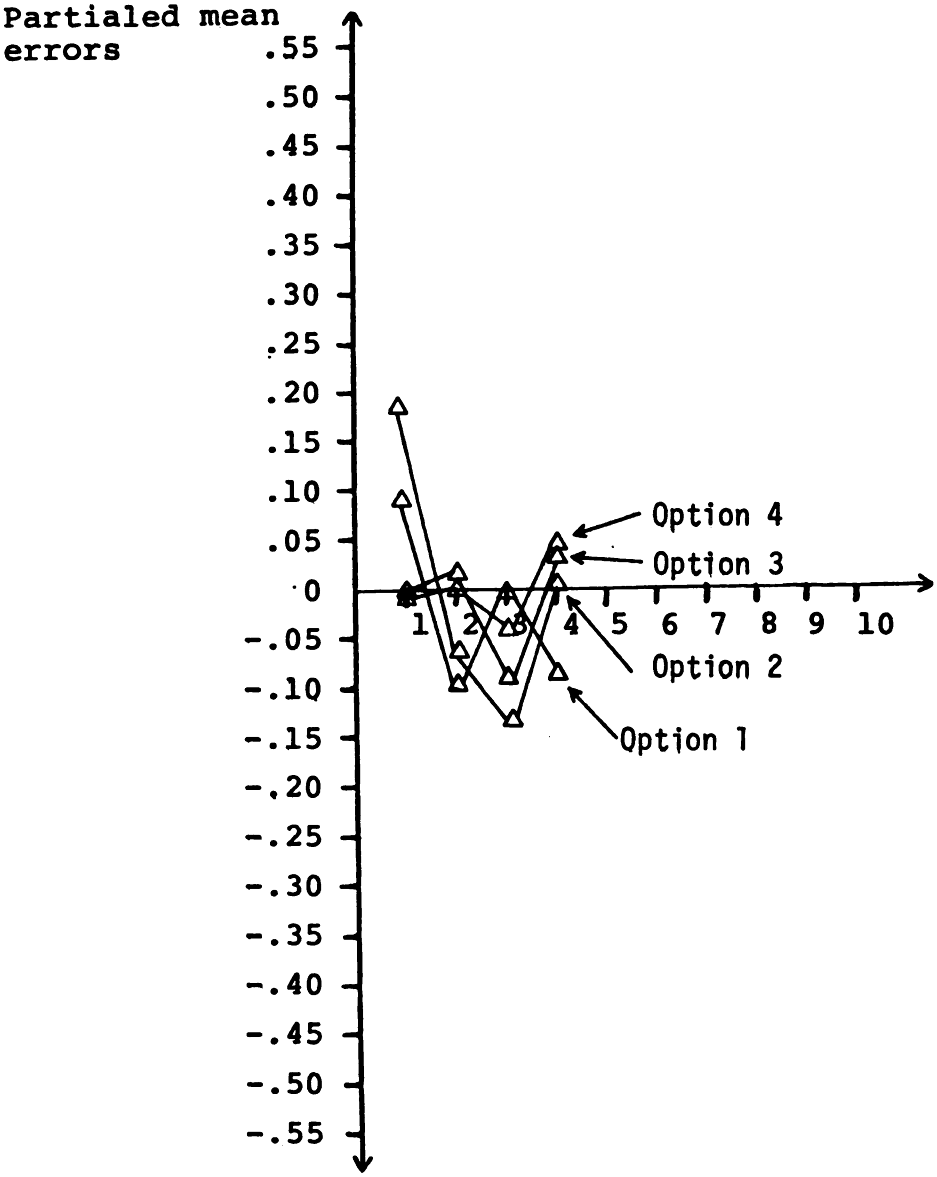


Figure 13

Subtest III: Option-Segregated Curves as a Function of Trial

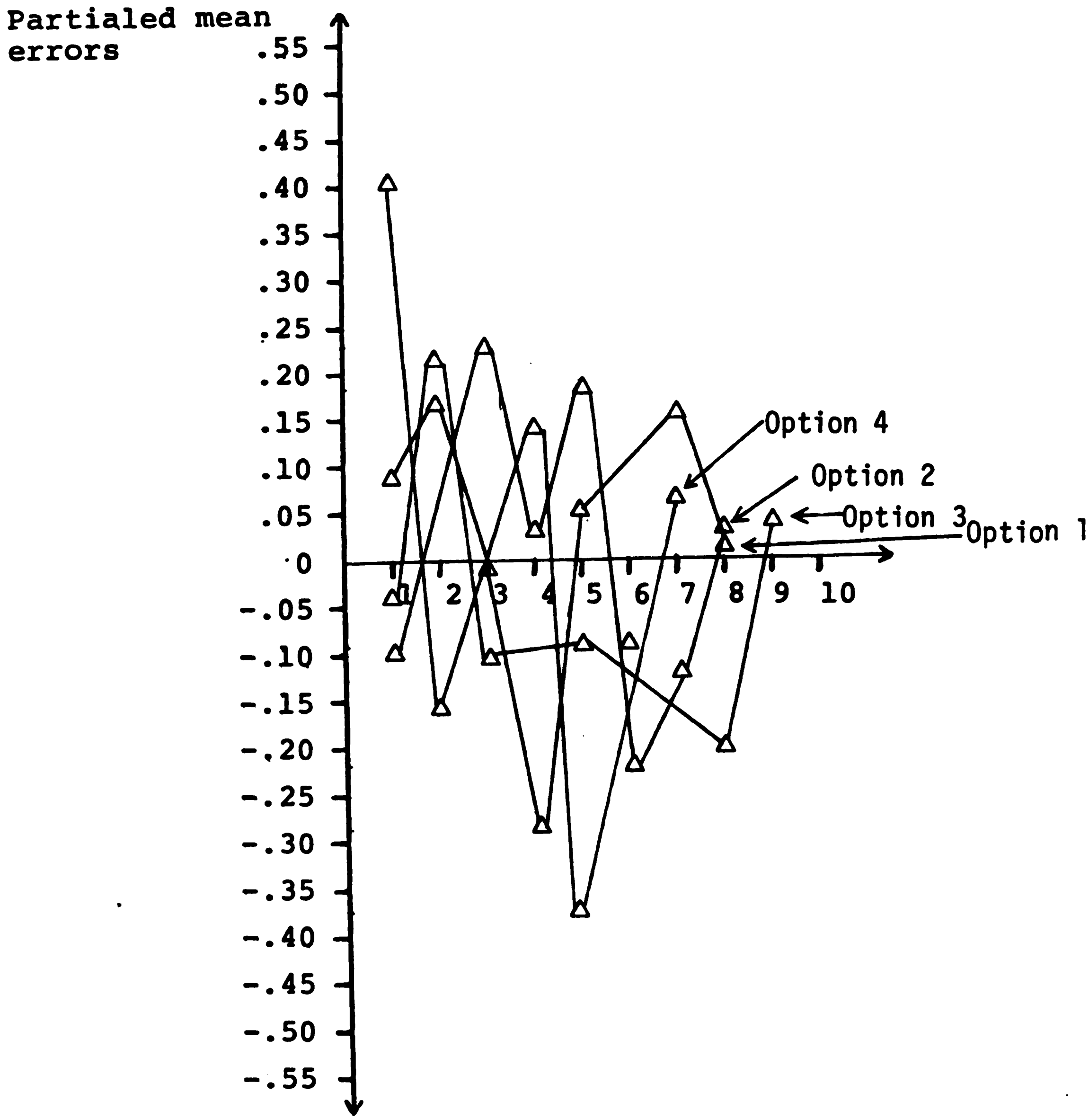


Figure 14

Subtest V: Option-Segregated Curves as a Function of Trial

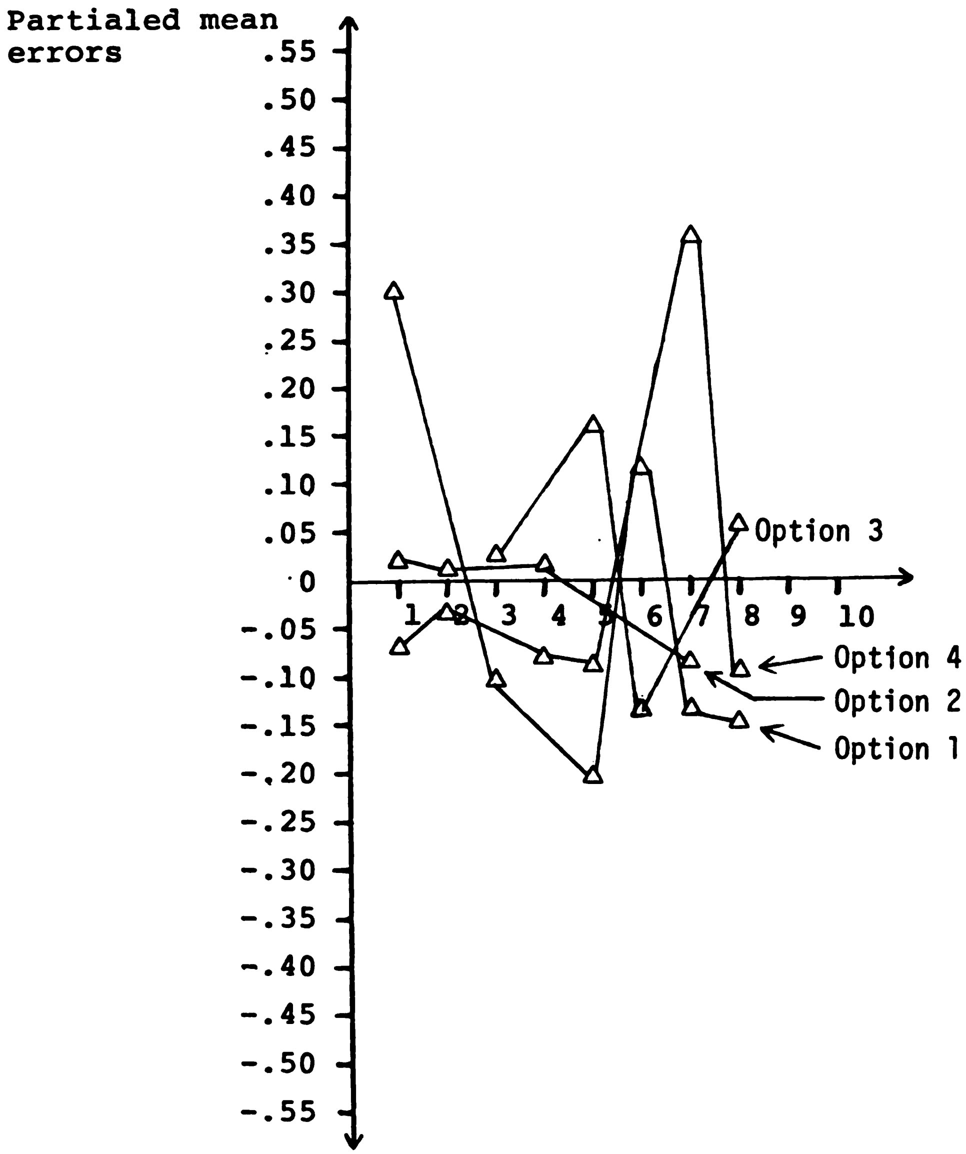


Figure 15

Subtest VI: Option-Segregated Curves as a Function of Trial

In order for the interactions to have obtained as predicted, it would have been essential for disparity in partialed mean errors, among options, to have decreased as a function of time, familiarity with the subtest, or the passage of trials. As became clear, to some limited extent this actually did occur, but the impact of sequence, or shifting stimulus characteristics, upon this interaction, was sufficiently large as to nearly overwhelm the predicted phenomenon. The results for subtest III served to illustrate this point very nicely.

Upon considering the results of the Option by Trial interaction for subtest III, it became apparent that the divergences in the curves appearing in Figure 13 were attributable, at least in very large part, to item, or sequence specific quirks in the design of the HCT. Thus, at trial 1, options 1 and 2 proved more difficult than options 3 and 4, and this pattern reversed (implying a clear interaction) at trial 2. Yet, an examination of Figure I-C in the Appendix disclosed that sequences 1, 2, 5, and 6 were over represented for options 1 and 2, trial 1, while se-quences 3, 4, and 7 were overrepresented for options 3 and 4, trial 1. On the other hand, at trial 2, sequences 3 and 7 were emphasized for options 1 and 2, while sequences 1 and 5 were emphasized for options 3 and 4. That the trial 2, options 3 and 4 curves did not rise so high as the trial 1, options 1 and 2 curves had, implied that learning (i.e., across trials) had obtained, but this could hardly be viewed or interpreted as an option by trial interaction.

At trial 3, the pattern shifted, and was characterized by greater divergence. Nonetheless, options 1 and 4 were experienced as more difficult than were options 2 and 3. Upon scanning Figure I-C, and Figure 1, it was determined that at trial 3, options 1 and 4 emphasized sequences 1, 2, and 5, which were the more difficult members of the set, while options 2 and 3 involved sequences 3, 6, and 8, which, on the whole, were less challenging for subjects.

Finally, at trial 4, options 2, 3, and 4 manifested increased mean errors, while only option 1 showed a decreased error frequency. From Figures I-C and 1 it was readily seen that at trial 4, option 1 involved sequences 3 and 8, which were characterized by reduced errors, while option 2 involved sequences 2 and 4, option 3 involved sequences 5 and 7, and option 4 consisted of sequences 1 and 6. In all three of these latter options, the sequences involved manifested larger mean error frequencies than did those characterizing option 1 at this trial.

In summary, the interaction between option and trial, though significant, as predicted,for subtest III, could not be clearly interpreted to demonstrate that familiarity with the test materially influenced the impact of option-defined response bias, over time. On the whole, in fact, the curves of Figure 13 did not appear to converge, over trials.

Essentially, the same was true of the option by trial interactions in both subtests V and VI, which are graphically represented in Figures 14 and 15, and whose partialed mean errors are summarized in Tables 25 and 26, respectively. In each of these cases, the essential character of the interaction was best clarified by appealing to sequence-specific shifts in item stimulus features. These findings were not summarized verbally because they added essentially nothing to what had already been reviewed concerning the impact and importance of the main effect for sequence.

It was noted, however, that in the cases of subtests V and VI the option curves indeed did converge as a function of trial. The convergence was not especially smooth or equally systematic at all places upon the curves, and this "noise" or distortion was attributed to the effect of sequence. Yet the points on the curves representing the four options were less discrepant and more coherently arrayed for the higher numbered than for the lower numbered trials, and this was essentially as had been predicted. It was concluded that within sequences, the effect of familiarity or experience with the subtests was to attenuate the existence of erroneous response sets or biases, as defined by disparities among partialed mean error scores for the various options.

Trial by Sequence

This interaction was predicted to emerge as significant, and it was believed that learning across trials would increase with time, or with the passage of sequence. For this to occur, it would be essential that sequence-segregated learning curves decrease in steepness, over time, or across sequences. Assuming an asymptotic learning function, performance, as defined by mean error scores, would then converge across trials toward some constant level of performance. The drop to the asymptote would then be steeper, by expectation, for the curves characterized by higher sequence numbers. The interaction was significant in the cases of subtests III and V, in which it accounted for .1% and .3% of the within-subjects variance, respectively. Results were summarized in Tables 27 and 28, and the associated curves were plotted in Figures 16 and 17.

For subtest III, and against the predictable background "noise" of sequence-determined and option-determined fluctuations in partialed mean errors, the results were roughly as predicted. Thus, the disparity in the curves at Trial 1 was somewhat greater than at subsequent trials, and particularly trial 4, indicating the asymptotic character of the curves. As well, the absolute decrease in partialed mean errors, over trials, tended to be larger for earlier than for later sequences, and this decrease in the slopes of the curves was consistent with learning, as was predicted. There was a slight tendency, as well, for later sequences, especially 6 and 7, to manifest more errors than might have been expected, during early trials. As illustrated in Figure 1, however, and commented on above, these sequences were defined by more complex stimuli than the others, and were consequently more difficult. The results of the trial by sequence interaction for subtest III led the current investigator to conclude that trial-to-trial learning had altered

Table 27   Subtest III: Cell and Marginal Means for Trial X Sequence

Sequence	Trial 1	2	3	4	Marginal
1	.341	.092	.144	.141	.179
2	.221	.109	-.034	.042	.085
3	-.034	.068	-.087	-.006	-.030
4	-.013	-.055	-.019	-.022	-.027
5	-.055	.036	-.116	.136	.000
6	.144	-.092	-.022	-.051	-.006
7	.090	-.195	-.053	-.081	-.062
8	-.024	-.147	-.236	-.158	-.142
Marginal	.076	-.023	-.054	.000	.000

Note. Entries are partialed mean error scores, as before.

Table 28. Subtest V: Cell and Marginal Means for Trial X Sequence

Sequence	Trial									
	1	2	3	4	5	6	7	8	9	Marginal
1	.162	.223	.099	.310	.180	-.003	.159	.034	.043	.134
2	-.038	-.065	-.121	-.004	-.085	-.172	-.122	.019		-.074
3	.405	.172	.057	.029	.053	-.153	.063	-.179		.056
4	-.149	-.331	-.262	-.281	-.372	-.287	-.283	-.214		-.272
5	.019	.506	.522	.124						.293
6	-.035	.003	-.061							-.031
Marginal	.060	.087	.040	.036	-.055	-.153	-.045	-.084	.043	

<u>Note</u>. Entries are partialed mean error scores, as before.

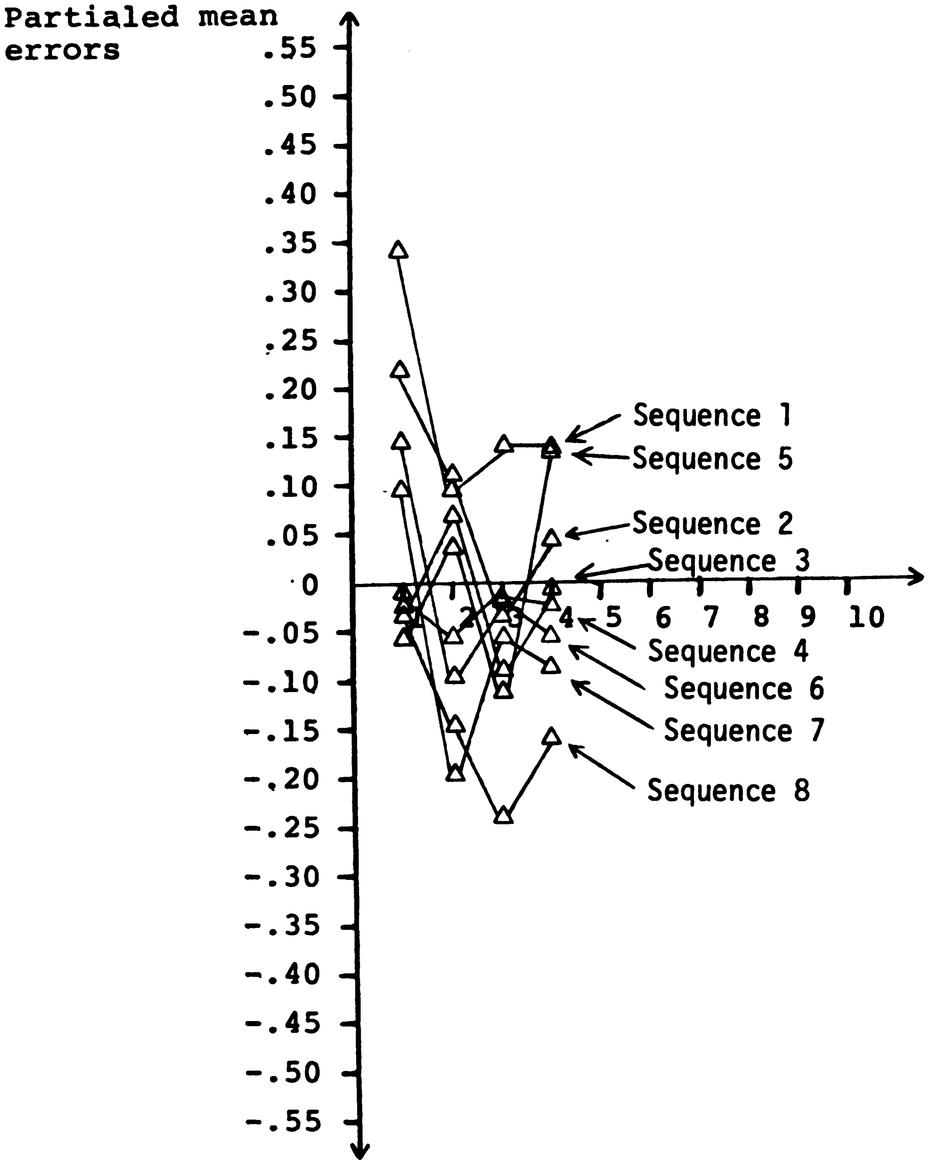


Figure 16

Subtest III: Sequence-Segregated Curves as a Function of Trial

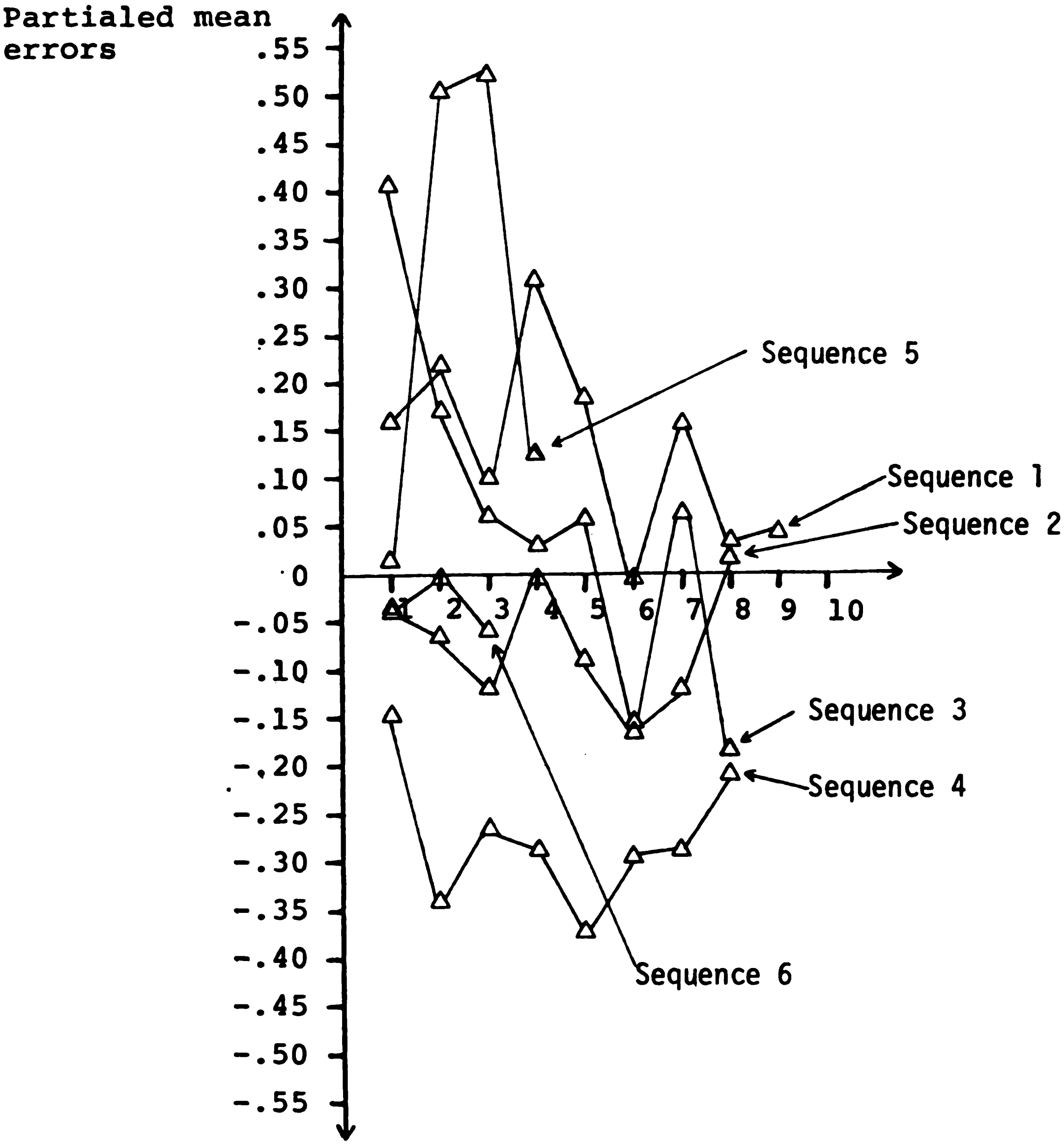


Figure 17

Subtest V: Sequence-Segregated Curves as a Function of Trial in the predicted way, as a function of sequence.

Much the same picture emerged from the consideration of subtest V, although, as was the case throughout, the findings were more complicated because of the wider discrepancy in sequence-specific item difficulties for this subtest. Thus, curves characterized by the higher numbered sequences manifested lower mean error scores for the earlier trials, and also demonstrated earlier asymptotes, than did curves defined by lower numbered sequences. Sequences number 3 and 5 were unusual, and from Figures 3 and I-E, it was readily concluded that this was attributable to their dramatically increased difficulty. It was inferred, again, that the trial by sequence interaction had occurred as predicted, for subtest V.

Damage

There was no plan to interpret this effect. It proved significant in every case, and as was summarized in Tables 9, 11, 13, and 15, it accounted for 1.4% of the within-subjects variance in subtest III, 13.4% in subtest IV, 3.1% in subtest V, and 5.6% in subtest VI. From a consideration of the valences of the regression weights, it was clear that Damage, as defined by the Augmented Impairment Index operated to predict HCT item behavior, as defined by partialed mean errors. Had these relationships been of particular interest to the current investigator, then doubtless they would have been hypothesized to emerge as in fact they did. However, the importance of brain damage, or rather its impact upon cortical impairment, was already discussed in the context of the between-subjects regression or variance analyses. In the present context, the variable, Damage, was applied merely so as to permit a partialing away of variance attributable to it in preparation for the analysis of interactions into which it figured prominently.

For each subtest, there were seven higher order interactions evaluated. Few of these proved themselves to be significant contributors to variance accounted for, within-subjects. For subtests III, IV, and VI, only one of these seven, or 14% of the total, emerged as significant. For subtest V, that manifesting by far the greatest within-subjects variance, two interactions, or 29%, were significant. By chance, 5% would have been anticipated to reach significance, and results obviously far exceed that expectation. Nonetheless, findings were concluded to have been very meager, at best. This was unsurprising, as, with the exception of the trial by sequence by option interaction, all the remaining combinations involved the variable, Damage, and consequently all were predicted to fail.

Sequence by Damage

This interaction reached significance for subtests V and VI. For subtests III and IV, it appeared to be entirely irrelevant. For subtest V, it accounted for .6% of the within-subjects variance, and for subtest VI, .8%.

To interpret these interactions, it was essential to examine the unique contributions of the component variables in the interaction sets. Statistics demanded in order to permit this were derived and summarized in tables 29 and 30, for subtests V and VI, respectively.

The significance of these interactions implied, essentially, that the relationships between brain damage, as defined by the Augmented Impairment Index, and HCT item performance as defined by partialed mean errors, shifted as a function of sequence. This was tentamount to the slope of the regression of HCT item performance upon brain damage shifting, dependent upon sequence, or item stimulus characteristics. Originally, although these interactions had been expected not to emerge significant, the competing hypothesis had been that the rate of learning, as defined by the reduction in partialed mean error occurring as sequence elapsed, would be greater for more cortically intact subjects than for more cortically impaired subjects. However, because sequence proved to be more intimately related to item stimuli than to the passage of time, this originally conceived alternate hypothesis proved useless and inapplicable.

As is hinted at by these remarks, the interpretation of the interaction between damage and sequence for subtest V bore no relationship to either the hypothesis the present investigator asserted or to its antithesis. Instead, and as was inferred from the quantities listed in Table 29, the differences in the relationship between brain damage and item errors were attributable to the relative difficulty of the items comprising the sequences of subtest V. Thus, for sequences of either extremely great or extremely small

Table 29. Subtest V: Summary of Sequence X Damage

Dummy	Unique R^2	D.F.	F	Significance	Slope[a]
1	.0040	1/5747	23.478	.001	.022
2	.0003	1/5747	7.140	.01	.047
3	.0002	1/5747	6.770	.01	.048
4	.0013	1/5747	21.865	.001	.023
5	.0003	1/5747	15.436	.001	.026

Note. Results were significant at or beyond the level specified.

[a]This is the slope of the regression of partialed mean errors upon augmented impairment index, with other variables "held constant", or, rather, statistically partialed away.

difficulty, slopes were about half the magnitude of those associated with sequences manifesting midrange difficulty. This was consistent with an argument to the effect that extremely difficult sequences of items were mastered only by the most cortically competent, while extremely simple sequences were failed only by the most cortically impaired. Sequences of midrange difficulty discriminated most effectively. This could be readily seen by comparing Table 29 with Figure 3. Another means of commenting upon the same result was merely to observe that for sequences of excessively difficult or excessively easy items, the variance and covariance quantities were sufficiently attenuated that regression slopes decreased (Magnusson, 1967, pp. 145-147).

For subtest VI, as could be seen by examining the values in Table 30, and considering the curve in Figure 4, the results were interpretatively somewhat different. In this case, the slope was over twice as large for sequence 1 than the sequence 2, when compared with the mean of the excluded sequences. In this case, then, it was thought appropriate to infer that for items earlier in the subtest, brain damage was more predictive of errors than for items later in the subtest, suggesting the possibility that learning was more difficult initially for those manifesting more cortical impairment than for those who were more cortically intact. Further along in the test, the distinction tended to fade somewhat indicating that asymptotic levels of learning were similar, irrespective of degree of cortical impairment.

Table 30. Subtest VI: Summary of Sequence X Damage

Dummy	Unique R^2	D.F.	F	Significance	Slope[a]
1	.004	1/3241	9.441	.005	.059
2	.004	1/3241	1.184	n.s.	.024

Note. Results were significant at or beyond the level specified.

[a]This is the slope of the regression of partialed mean errors upon augmented impairment index, with other effects having been partialed out of relevance.

Trial by Damage

Only for subtest III did this interaction prove even marginally important, and for this subtest, it accounted for some .4% of the within-subjects variance. To accomplish interpretation of the interaction the quantities summarized in Table 31 were computed.

On the basis of the values in Table 31, considered in conjunction with results depicted by Figure 5, it was inferred that performance at later trials, as opposed to earlier ones, was more substantially correlated with the Augmented Impairment Index. Substantively, this implied that learning was more readily acquired, the less cortical impairment existed. Performance at early trials did not discriminate well between those with and without brain damage, but across trials, those without cortical impairment tended to manifest a greater decrease in errors than those with cortical impairment. This result in effect supports the notion that the learning process itself occurred as a function of the degree of cortical impairment present. This finding was consistent with the notion that a qualitative distinction exists between the performance of brain damaged and intact subjects.

Trial by Sequence by Option

This interaction was expected to reach significance, but it did so only for subtest IV, and in which it accounted for .7% of the within-subjects variance. Only a small number of the total 28 component variables for this

Table 31. Subtest VI: Summary of Trial X Damage

Dummy	Unique R^2	D.F.	F	Significance	Slope[a]
1	.003	1/4768	15.431	.001	.017
2	.000(2)	1/4768	7.119	.01	.028
3	.001	1/4768	2.943	n.s.	.037

Note. Results were significant at or beyond the level specified.

[a]This is the slope of the regression of partialed mean errors upon augmented impairment index, with other effects having been partialed away.

interaction actually contributed materially to its significance. In order to simplify interpretation, only those members of the interaction set which themselves accounted for significant variance were considered. These variables, with associated quantities essential for comprehending the interaction, were summarized in Table 32.

In general, as can be seen by persuing the tabled results, the variables, or individual cells, which emerged as significant, favored on options 3 and 1 (the hardest, and easiest, respectively), sequences 2, 3, and 4 (those early in the subtest, and following the "instructional" sequence), and trials 1 and 4 (again, the most and least difficult, respectively). The first, fourth, and sixth members of the set included in the table were all for too difficult to be accounted for by either the sequence or the trial. All three of these were characterized by option 3, the most difficult of the lot, by far, and these results could thus be interpreted as the influence of Option, as an indicant of the erroneous response bias, on the part of subjects. It will be recalled that the problem with options 3 and 4, subtest IV, was that they were almost invariably confused, by transposition. Items calling for either of these as the correct response were missed in over one half of the cases. The variables mentioned which reached significance did so because in other respects (i.e., trial number, sequence number) the expectation was that error frequencies would have been lower, save that the occurrence of option dramatically inflated these. The fifth member of the set

Table 32. Subtest IV: Summary of Trial X Sequence X Option

O[a]	S[b]	T[c]	Mean[d]	E(S)[e]	E(T)[f]	Unique R^2	D.F.	F	Significance[g]
3	2	3	.061	.066	.017	.0032	1/5813	15.637	.001
1	2	1	-.068	.066	.033	.0007	1/5813	14.814	.001
2	2	4	.055	.066	-.001	.0001	1/5813	7.116	.01
3	3	7	.173	.131	.033	.0010	1/5813	6.971	.01
3	4	1	-.032	-.007	.033	.0001	1/5813	6.636	.01
3	9	1	.168	.065	.033	.0002	1/5813	4.484	.05
1	4	4	-.078	-.007	-.035	.0005	1/5813	4.062	.05

[a]Option.

[b]Sequence.

[c]Trial.

[d]Cell mean, defined, as conventially, as partialed mean errors.

[e]Mean for associated sequence level.

[f]Mean for associated trial level.

[g]Results were significant at or beyond the level specified.

manifested a lower error frequency than expected, both given its associated trial and sequence levels, and especially given its option level, that of 3. Examination of Figure I-D clarified what had occurred: the stimuli in sequence 4 were all somewhat engineered to elicit "three" responses from subjects, because they all manifest three distinct segments. Thus, the errors for the cell in question, item number 14, were attenuated, due to the intercession of an inappropriate response bias resulting in fallaciously correct answers.

The second and seventh elements of the set summarized in Table 32 were both far easier than expected, upon the basis of Trial and Sequence marginal means. The reason for this was that they both featured option 1, by far the easiest of the lot, both because it, like option 2, was far less often transposed, and as well because subjects frequently selected it, apparently as a result of their having conceptualized item stimuli as unitary events, evoking the association, "one".

The remaining member of the set, number three, could not be interpreted on the basis of its option number. Option number 2 was relatively easy, and for trial 4, this item was far too difficult. Interestingly, the associated sequence number was 2, and this suggested that this variable alone could be unequivocally interpreted as supportive of the notion that the disruptive impact early on of inaccurate response sets, as assessed by response option, was modified favorably by familiarity with the subjects' organizing principle, as taught by positive reinforcement and punishment. Thus, this item was inferred to have been difficult beyond expectation because it occurred early in the subtest.

Trial by Sequence by Damage

This was the last interacton remaining to be interpreted. It emerged as significant only for subtest V, and accounted there for 1.0% of the within-subjects variance. Its significance was unexpected, originally. Relevant quantities, essential for the interpretation of this interaction, were summarized in Table 33.

The appropriate interpretation for those particular trial by sequence combinations, or items, which were significant was that the individual slopes of the regression lines for partialed mean errors upon Augmented Impairment Index were unequal; i.e., that homoscedasticity did not, in fact, obtain. Another means of interpreting this same point was that the significant items were either more discriminative or less discriminative of brain damage, as a function of mean errors, than would have otherwise been expected.

From the slopes of the regression lines, it was readily inferred that in most cases the items were less, or rather, inversely discriminative of brain damage. Rather, then, the prediction equations which involved only brain damage, other main effects, and lower order interactons had overemphasized the salience of the relationship between impairment and item performance, for these specific items. The exception was that for the fourth member of the set, item number 24 for

Table 33. Subtest V: Summary of Trial X Sequence X Damage

S[a]	T[b]	Unique R^2	1/5710	41.334	Significance[c]	Slope[d]
1	2	.0001	1/5710	4.629	.05	-.094
1	5	.0011	1/5710	3.975	.05	-.053
2	3	.0000	1/5710	4.324	.05	-.055
3	7	.0011	1/5710	4.533	.05	.045
3	8	.0006	1/5710	11.334	.005	-.020
4	4	.0007	1/5710	3.784	n.s.	-.076
4	8	.0001	1/5710	5.514	.05	-.045
5	7	.0002	1/5710	3.194	n.s.	.002

[a]Sequence.

[b]Trial.

[c]Results were significant at or beyond the level specified.

[d]This is the slope of the regression of partialed mean errors upon the augmented impairment index, for the cell, as indicated. Other effects have been partialed away.

subtest V, the relationship had been underestimated.

Although the items were considered in some detail, with reference to their fates vis-a-vis the main effects for item and sequence, their roles in previous interactions, and their stimulus features, it proved impossible to delimit common characteristics possibly accounting for their significance. It was noted, however, and as was commented upon previously, that subtest V manifested, throughout the analyses, more item variance than the other subtests. It was previously inferred, as well, that this was attributable to the interference of the response bias acquired during subtest IV, and, as well, to the dramatic changes in item stimuli, between sequences, as the test progressed. Perhaps it was the case that the substantial fluctuations in error rates imparted by these factors were responsible for the significance of the present interaction. Certainly, it was in no way plausible to infer the operation of a systematic shift in trial by sequence *learning*, as a function of impairment, on the basis of the results of this interaction.

Remaining Interactions

The Option by Damage, Sequence by Option by Damage, and Trial by Option by Damage interactions uniformly failed to reach significance. This had been expected, because of Reitan's (1958, 1959) previous arguments that the process of learning was no different for the cortically impaired than for the cortically intact.

Discussion

With the attendant qualifications stipulated and elaborated below, findings of the present investigation indicated that learning, defined as a decrease in errors over time, or as a function of familiar-ity with a given subtest, could be shown to occur in HCT item responses.

The first pertinent qualification formulated was that learning, as defined herein, was a determinant in this regard of modest importance. Though it accounted for sufficient variance in subjects' responses to attain significance, learning was not, that is to say, an especially salient predictor of HCT item errors.

The second relevant qualification derived was that two features seemingly intrinsic to the HCT itself were far more dramatically important predictors of HCT item performance than was learning. These included item stimulus characteristics and, essentially, an interaction between subjects' response biases and the characteristics of the item stimuli. The first of these successfully competing factors was represented by the effect, 'Sequence' and as well, a number of the interactions involving Sequence. The second of these factors was represented by the effect, 'Option', and several of <u>its</u> interactions.

Both of these effects, had been considered by the present investigator, but the importance of the sequence-specific item stimuli features proved far greater than anticipated, and the impact of Option, though expected, was falsely assumed to have been "removed" by entering this effect <u>first</u> in the regression analyses. The failure of this factor to ameliorate the problem was readily traced to the nonrandom and (by implication) incomplete crossing of Option with Trial and Sequence.

To briefly review the results, essentially the same nonrandom frequencies and distributions of errors across options were found in the present investigation as has been documented previously by Simmel and Counts (1957). As they emphasized, errors generally occurred when subjects attempted to apply response sets, or organizing principles, acquired on the basis of experience with preceding subtests. Also relevant was that when subjects were unclear about the correct organizing principle or response set, then they tended to resort to counting some aspect or aspects of the item stimuli, and to utilize this counting as a basis for deriving a response. Counting might, as Simmel and Counts (1957) argued, have arisen as a "last alternative" in a situation precluding the clear application of any other strategy. As well, counting was learned by experience with subtest II, and may have merely reappeared in situations of vagueness. In any event, the appearance of responses biases resulted in nonrandom distributions of errors across response options, and certain items and/or subtests tended, as discussed, to elicit certain erroneous response sets, in systematic ways. This systematicity was removed, by entering the effect, 'Option' first.

For each of the subtests, the factors, 'Sequence' and 'Trial', proved significant, as predicted. The values summarized in Table 18 and depicted in Figures 1 through 4 indicated that errors indeed tended to decrease, across sequences. The same result was noted across trials, as shown by Table 19 and Figures 5 through 8.

Findings tended to be somewhat more complex than had been anticipated, and this was emphasized by a consideration of the orthogonal polynomial trend components. Pertinent results were included in Tables 10, 12, 14, and 16, for subtests III, IV, V, and VI, respectively. As expected, the trend components were significant. However, the curvilinearity apparent was generally traced either to nonrandom pairings of levels of Trial and Option, or to sudden shifts in item difficulty between levels of sequence, as effected by alterations in the features, and presumably the complexity, of the item stimuli.

Thus, for subtest III, although linear, quadratic, and cubic components emerged as significant for the generic effect, 'Sequence', the linear approximation proved the better. For this subtest, both the quadratic and the cubic components were likely attributable to the shift in item difficulty between sequences 4 and 5, as item composition suddenly became more complex. This increase in complexity was then accommodated to gradually by subjects, and items in sequences 7 and 8 were again experienced as progressively less difficult.

For subtests IV, and V, the picture was muddied even more, and this also was traced to marked alterations in item stimulus composition between sequences. These grossly irregular curves, depicted in Figures 2 and 3, manifested large quadratic and/or cubic components, but they were hardly of the predicted exponential and asymptotic form. Although simpler, subtest VI manifested similar problems.

Insofar as the effect, 'Sequence', was concerned, then, it proved impossible to declare, with an acceptable degree of certainty, that the curvilinear character of the curves had anything whatsoever to do with learning rather than with shifts in item difficulty which were designed into each of the subtests. Halstead clearly increased the difficulty of items as a function of Sequence, within subtests, and this proved hopelessly confounded with the competing process of learning, also occurring as a function of Sequence, within subtests. Because of this difficulty, curvilinear components were not further considered.

The linear aspects of the Sequence curves tended to fare somewhat better, with the exception of subtest V, in which average errors actually <u>increased</u> as a function of Sequence, in direct opposition to what was predicted. Slopes for the sequence curves for subtests III, IV, and VI, that is, were negative, and it was inferred that learning indeed did occur, within subtests and across sequences, as familiarity with the organizing principles developed in subjects.

As was commented upon concerning Sequence, although the picture which emerged proved more complex than had originally been anticipated, genuine and recognizable learning curves obtained when mean errors were graphed as a function of Trial. Again, because of certain features designed into the HCT, and due as well to the rather unbalanced character of the design, results were not as clear cut as might otherwise have been the case. For Trial, as was indicated above, the interfering factor was the nonrandom pairing of levels of Option with levels of Trial. Nonetheless, upon restricting consideration solely to the linear aspects of the trial curves, it was concluded that the presence and operation of learning had been demonstrated.

It had originally been predicted as well, that both Option by Sequence and Option by Trial interactions would prove significant, and moreover, that inspection of the cell means would reveal that the disorganizing effect of Option would decrease across levels of Sequence or Trial, again illustrating the effect of learning. Likewise, for the Trial by Sequence interaction, it was anticipated that the across-trials decrease in errors would accelerate as a function of Sequence. These hypotheses failed to receive unequivocal support, and again, this failure was traced inevitably to the impact of Option and Sequence, as reflective of subjects' response biases and HCT item characteristics, respectively. In some cases, the Option by Trial interactions indeed did obtain as predicted, as the Option-segregated curves across Trials did indeed converge in level of experienced difficulty. This finding did not obtain for Sequence, and it was inferred that Sequence admitted significant influence in the form of rather dramatically shifting item or stimulus characteristics, and that this operated in turn to obliterate the effect of familiarity with the subtests which might otherwise have accumulated.

For the Trial by Sequence interactions, it was also fairly readily determined that the salient effects were attributable to nonrandom sorting of options across trials.

The Sequence by Damage interaction proved significant for both subtest, V and VI, but for subtest V this was again related to item difficulty, as defined by the stimulus characteristics featured by the various sequences. It was noted that at either extreme, too difficult or too simple, the subtest tended somewhat to lose its capacity to discriminate or predict brain damage, while this was preserved for sequences of midrange difficulty or complexity. For subtest VI, the picture which emerged actually supported the notion that learning early during the subtest was more dependent upon cortically intact functioning than learning later in the subtest. In essence, this suggested that the hypotheses of Goldstein (1940, 1942) may be at least partially accurate, irrespective of Reitan's (1958, 1959) arguments and empirical findings to the contrary.

The significant Trial by Damage interaction for subtest III revealed that for early trials, across sequences, extent of cortical impairment was less relevant a determinant of errors than for later trials, implying that cortically intact subjects were better able to utilize the feedback, or reinforcement, they received, in the interest of altering their response sets in the correct direction, and thus decreasing errors. They were, that is to say, better able to learn from their mistakes and to divine the correct organizing principle. The finding was weak, and was apparent only for subtest III.

The sole significant example of the Trial by Sequence by Option interaction, that involving subtest V, was also quite easily interpreted as the operation of the usual nonrandom distribution of elements in the product set, Trial X Sequence X Option. Because, that is, of systematic differences arising as a function of sequence-specific item characteristics, and information concerning response biases conveyed by option, the incompletely crossed factors reached significance. In no sense did this occur for the reason expected, that of the across-trials decrease in interference by incorrect response sets accelerating as a function of sequence.

The Trial by Sequence by Damage interaction involving subtest V, also significant, was similarly interpreted.

Perhaps the most succinct and appropriate summarizing comment the present investigator reached was that the hypotheses as outlined were not given an adequate appraisal. Again, this difficulty was attributable to three problems with the design, or more accurately, with the HCT itself.

The first major flaw in this regard was that sequence emerged as a hopelessly confounded mixture of the effects of time and familiarity with a given subtests' organizing principle, and of dramatically shifting item stimulus composition. Because of this confounding, it was often not possible to assess or ascertain the relative contributions to error frequencies of learning and item difficulty, or complexity. Essentially, this meant that 'Sequence', as conceptualized by the present investigator, failed to exist. Item characteristics shifted sufficiently between sequences that in many cases (e.g., subtest V) the effect was tantamount to beginning another subtest, with a new and distinct organizing principle. It often seemed that Ward Halstead designed into the HCT every characteristic he could, within reason, which would challenge and disrupt the operation of learning.

The second salient problem involved the nonrandom sorting of option levels across trial levels, as this confounded the operation of response bias with that of learning, across trials. Again, this prevented the drawing .pa of clear inferences concerning fluctuations in error frequencies for the levels of Trial.

Finally, the factors in the design, Option, Trial, and Sequence, were incompletely crossed. This flaw, along with the difficulties already mentioned concerning Sequence and option, led to a number of significant interactions which meant nothing more startling than the usual sequence-specific and option-specific effects upon HCT item behavior.

Depending upon one's theoretical predilections, these problems can be viewed as unacceptable flaws in the experimental design, or as unavoidable consequences of attempting to evaluate an in vivo situation. From the point of view of the classical variance analysts, it is the experimenter's first responsibility to guarantee that correlated effects do not exist, by carefully balancing and completely crossing the factors in her or his design. From the perspective of more contemporary (e.g., Kerlinger and Pedhazur, 1973, pp. 29-52) thinkers, however, it is acknowledged that sometimes effects are correlated, and to demand otherwise is to distort reality such that conclusions emerging are not generalizable.

The HCT does not represent a balanced design, nor are certain factors (e.g., item difficulty) well controlled. But to manipulate its structure so as to assure balance and control would be to change the measure sufficiently to render the product incomparable to the measure as it currently exists. Consequently, it is felt appropriate to assert two points. The first of these is that a rigorous examination of the factors as outlined herein, Option, Sequence, and Trial, would absolutely demand the alteration of the HCT in ways permitting the full and orthogonal crossing of all factors involved. As well, item difficulty would have to be more carefully managed, across sequences. The orthogonal crossing of Option with the other factors, is particularly, important. Such a carefully wrought design would permit an adequate scrutiny of the hypotheses preferred by the present investigator.

The second point to be made here is that even with the HCT itself, it proved possible to demonstrate that learning, defined as reductions in errors over time, and with familiarity with the HCT subtests, is a relevant determinant of

HCT item performance. Given the design of the HCT, the importance of learning was modest, or at least the extent to which it could be unequivocally shown to obtain and operate was modest. At present, then, it is reasonable to conclude that the HCT nicely predicts brain damage, not learning.

Author's Notes

[1]Today, we know this, essentially, as the principles of discrimination (non-equivalence) and generalization (equivalence).

[2]In contradistinction to the current generation of learning theorists, Kluver and others of his ilk, sometimes called "structuralists" or "mentalists", had moved, so to speak, the study of human (and infrahuman) behavior back into the organism (Rachlin, 1976, pp. 35-36). It was felt that the organism operated upon, or transformed information, and moreover, that these "operations" were substantially amenable to the full rigor of experimental study (Kluver, 1925).

[3]A couple of interesting digressions may be inserted here. First, and consistent with most others of the same period, Goldstein (1936, 1939, 1944) believed the frontal lobes to be largely responsible for the personality, intelligence, and, most importantly, the abstract attitude. Actually, much later, Luria (1973) argues a similar point, attributing self reflection, the control of impulses, and behavioral planning, programming, and execution to the frontal areas. It is also fascinating to reflect upon the essential

identity between operations Goldstein included in his concept, 'abstract attitude', and those Hartmann (1958) was referring to with *his* concept, the conflict-free, or *executive*, functions of the ego. Finally, Halstead (1947, 1948, 1951) himself believed the frontal lobes were directly or indirectly responsible for planning, organizing, executing, and, as it were, detachedly appraising what he called *adaptive behavior* (Halstead et al., 1946), or *Biological Intelligence* (Halstead, 1951). The similarity of these constructs is striking.

[4]Halstead was heavily influenced by a longstanding tendency (Walsh, 1978, pp. 113 ff.) to champion the frontal lobes as the seat of all that was viewed as essentially human (including intelligence, problem-solving behavior, and so on). To his credit, of course, Halstead (e.g., 1947) invariably provided careful measurements of location and magnitude of cortical lesions, and found repeatedly that his assumptions about frontal hegemony were empirically borne out. Fairly recently, Chapman and Wolff (1959) have reanalyzed much of Halstead's original data after having adjusted his measurement procedures, which apparently tended to overestimate the magnitude of frontal lesions. Their finding was that lesion location exerted no impact upon grouping behavior or any other complex activity, but rather that only lesion size was significant.

[5]The original version is very thoroughly described in Halstead and Settlage (1943). The contemporary version (or

versions, actually, as forms exist for young children, older children, and adults) is even more explicitly typified in Reitan and Davison (1974). A thorough conceptual and pictorial analysis of the current version also is to be found in the fine, if obsessive, monograph of Simmel and Counts (1957). Finally, the Method Section of the current endeavor also describes the current adult version of the procedure in fair detail.

I frequently call items "trials," and invariably think of them in this way. Basically, this is because items are not experimentally independent events, and, moreover, are formally similar to one another, though this varies. Items, consequently, are nicely conceptualized as repeated measures (i.e., of abstracting ability, problem solving behavior, brain damage, or whatever else) to which the respondent is subjected.

[6]Hanfmann and Kasanin (1937) actually came close to Halstead's rigor with their object sorting procedure, yet the method of reinforcement is largely verbal, and thus is not so standard as Halstead's. A more recent, and currently popular approach is the Wisconsin Card Sorting Test (Milner, 1963). Here again, feedback is verbal, though formally the learning aspect of the test can be though of as closely resembling the HCT. Both of these alternate procedures are cheaper, faster, and less unpleasant to subject and administrator than the HCT. Yet both are formally simpler tasks, demand less complex cognitive activity of the subject, and

rely more heavily upon the examiner for uniformity of administration.

[7]Halstead (1947) also retained the ten most sensitive or discriminating tests with regard to brain damage (presence or absence) and with these derived his "Impairment Index," the proportion of these indicators scoring in a given case in the brain-damaged range. This set of ten indicants overlapped significantly with the refined set of 13 indicators factor-analyzed by Halstead (1947). It also was later examined by Reitan (1955) in a now classic study of the validity and diagnostic efficiency of these ten "best" indicators. Results of that study led Reitan (1966) to discard three of the variables, including: (1) Critical Flicker Frequency, (2) Critical Flicker Frequency - Deviation; and (3) Time Sense Test - Memory Component, because they discriminated poorly between matched cortically impaired (*n* = 50) and cortically intact (*n* = 50) groups. Both subsets of Halstead's (1947) contained the Category Test, as did Reitan's (1955).

[8]Actually, "arational" would have been more accurate than "irrational", for it was, according to Halstead (and after Goldstein) characterized by a lack of detached, abstracting, and probably goal-directed cognitive activity.

[9]Personal communication with Ms. Dorothy Ford, previously secretary to Ward Halstead, and still employed by the University of Chicago.

[10]Efficiency is here conceptualized as by Meehl and Rosen (1955), viz., the likelihood of false negatives and false positives, with recourse to the base rates of the relevant dichotomous event, here brain damage vs. no brain damage.

[11]Values are negative because the authors transformed raw scores to T-scores (mean 50, standard deviation 10), with higher scores indicating <u>better</u> performance. This contrasts with standard procedure (Halstead, 1947), which involves a sum of errors, and consequently, higher scores consistent with poorer performance.

[12]Actually, the complete picture in this context is more muddy and complex than this, and there are at least strong suggestions (see Malec, 1978) that the Halstead-Reitan Battery <u>is</u> useful in distinguishing between schizophrenics and the cortically impaired. Lenzer (1980) has provided a reasonably incisive and provocative review, with the suggestions that Halstead's tests may be sensitive, not to brain damage per se, but rather to the erosion of the cognitive processes, and that 'schizophrenia' is tantamount to some (as yet indeterminate) subtle form of brain damage, or at very least, dysfunction. In this latter regard, Tucker (1980) has summarized an exciting corpus of studies, and offers truly interesting inferences concerning at very least the neurochemical and neuropsychological aspects of schizophrenia and the affective psychoses. Reitan (1974) himself has acknowledged that psychiatric-neuropsychological differential diagnoses remain subjective.

From the perspective of any serious study of psychopathology it would seem apparent that we are significantly (not entirely) confused about what it entails, how to differentiate one condition from another, etiological and prognostic accounts, and so forth. It seems highly likely that the capacity to distinguish between cortical impairment and the "functional disorders" will coincide with a good neurobehavioral account of the latter. This, however, is not shortly forthcoming.

[13]Actually, the majority of Reitan's many published studies during the latter half of the 1950's utilized the same 100 subjects, divided into matched pairs of brain damaged and intact individuals.

[14]Shaw, with a diverse sample of 674 neuropsychological referrals, obtained an odd-even split-half reliability coefficiency (i.e., of equivalence) of .98, and a standard error of measurement equal to 4.47, relative to raw error score. Interpretation of this quantity remains moot, as the items are in no sense independent events. The author felt he was avoiding this problem by utilizing an odd-even split, but of course this is at least debatable, and probably frankly erroneous.

With several different samples, Matarazzo et al. (1976) have also obtained test-retest reliability estimates of .60 (*n* = 29, young normal males), .72 (*n* = 35, chronic schizophrenic patients), .82, (*n* = 15, carotid endarterectomy patients), and .96 (*n* = 16, patients manifesting diffuse cerebrovascular disease), with 20 week interims. These quantities, of course, are less psychometrically moot than Shaw's (1966).

[16]I frequently *call* items "trials", and invariably think of them in this way. Basically, this is because items are not experimentally independent, but are formally identical with one another, and consequently are nicely conceptualized as repeated measures (i.e., of abstracting ability, problem solving, brain damage, or whatever).

[17]As Cronbach (1953) has pointed out, the correlation effectively nullifies differences in paired scores due to elevation (mean differences) or scatter (differences in variance). Mathematically, this is due to its being based upon paired mean *deviation* scores standardized by their being interpreted *relative to* their standard deviations. The interested reader is directed to a series of publications by Cronbach (1953, 1955, 1958) and Cronbach and Gleser (1953).

[18]It is interesting (if a bit distracting) to point out that this is precisely what Halstead (1940, 1943) was interested in when he began the sequence of studies that ultimately led to his developing the HCT. In fact, prior to labeling his approach a study of "categories," he thought of it, after Kluver (1929, 1931, 1936), as an application of the "method of equivalent stimuli" to the problem of human learning. Thus, his concern was with developing an understanding of the process or processes by means of which two stimulus

arrays will or will not elicit identical responses. Brunswik (1956) came at the problem a little differently, in that his emphasis was upon delimiting those <u>aspects</u> of a stimulus which the processing organism abstracts, as it were, in the activity of formulating and emitting a response. Brunswik also became far more intrigued by perceptual operations on the stimulus array (Hammond, 1966), while Halstead clearly concerned himself with the "reasoning" (i.e., abstraction) intervening between stimulus and response. Of course, Halstead's definition of stimulus and response were far more restrictive than Brunswik's.

[19]Simplistically, I define the usual, non-speeded test of capacity (e.g. WAIS information or comprehension subtests) as, more or less, a cross-section appraisal of, say, intelligence, at a given point in time. Though <u>some</u> learning may occur while a subject is taking such tests, it is likely to be minimal, and of minimal importance insofar as its impact on total score is concerned. Thus, the cognitive "picture" one gets of the individual is a static one. The HCT, on the other hand, evaluates the learning <u>process</u> and not just its product. Both are evaluated, of course, and this is witnessed by the correlations between HCT and WAIS Intelligence estimates (Reitan, 1956). This, to me, implies that the test is best viewed as a longitudinal section, or dynamic view, of capacity, or the appraisal of capacity as process.

[20]The investigators eliminated subtests I and II, because, as Halstead (1943) emphasized, these are included solely to

provide practical instructions to the respondent, and items within these subtests are rarely incorrectly responded to. Subtest VII was omitted, of course, because it contains a mixture of items from all of the previous subtests, and is a test of memory, more or less.

[21]The computation of alpha, and for that matter, the product-moment correlation coefficient, depends upon items being experimentally independent trials, and this assumption is almost certainly violated by the HCT. Interpretation of these values, thus, must remain moot, at least for the time being.

[22]Indeed, from Halstead's (1943) original version, to Reitan's (1955) contemporary version of the HCT, subtest III has consistently proven more difficult than the remainder of the subtests.

[23]Of course, the requirement of memory is implicit in the solution of _each_ of the subtests.

[24]I am aware that memory is a component of (i.e., perhaps necessary, but not sufficient for) learning. However, for purposes herein, learning is defined as the increment, over time, in the probability of responding correctly to items.

[25]In the usual design without repeated measures, 'experimental units' are just subjects. Here, they are subjects x items.

[26]Essentially, MLR, (ANOVA, too) permits two, alternate treatments of error: incremental and fixed as small as possible. I have used a combination of the two, more or less.

[27]It will be recalled that HCT total scores figures into the seven-indicator composite, Impairment Index.

[28]Technically, 'Trial' is nested within 'Sequence', implying that trials one through four, sequence one, are not equivalent to trials one through four, sequence two, because of intrinsic (i.e., trial-dependent) differences. I make the assumption that trials <u>are</u> identical, and that the differences which <u>do</u> emerge will be explicitly attributable to the other factors in the design, which will already have been partialed away prior to entering the interaction, 'Trial, x Sequence', or analogously, the main effect, 'Trial'.

APPENDIX

APPENDIX

The Halstead Category Test

It would certainly be possible to produce yet another original description of the Halstead Category Test (HCT), but it would, unfortunately, prove patently impossible to improve upon the better, existing ones. Consequently, these will be relied upon. A good general description of the character and materials of the HCT is as follows:

> This test utilizes a projection apparatus for presentation of 208 stumulus figures on a milk-glass screen. An answer panel for use by the subject is attached to the test apparatus and is located at a convenient level below the screen. The answer panel contains four levers which are numbered from 1 to 4. The subject is told that he should inspect each stimulus figure when it appears on the screen and depress one of the four levers, depending upon which answer he thinks may be correct. Depression of any of these levers will cause either a bell or a buzzer to sound depending upon whether or not the lever selected is the "right" or "wrong" answer. Only one response is allowed for each item. Before the test begins, the subject is told that the test is divided into seven groups of pictures and that each group has a single principle running through the entire group from beginning to end. On the first item in any group, the subject can only attempt to guess the right answer, but as he progresses through the items of the group, the sound of the bell or buzzer with each response indicates whether his guesses are correct or incorrect. In this way, the test procedure permits the subject to test one possible principle after another until an hypothesis is hit upon which is positively reinforced consistently by the bell. The subject is never told the principle for any group regardless of the difficulty he might

> encounter, but the first and second groups are nearly always easily performed even by persons with serious brain lesions.
>
> (Reitan and Davison, 1974, pp. 366-368)

The subtests are depicted below in Figures I-A through I-G (i.e., for subtests 1 through 7, respectively). It bears mentioning that in these seven tables, the letters appearing sporadically beneath the stimuli indicate that the stimuli are colored, and represent the hues, red (R), blue (B), green (G), and white (W). Unless so indicated, stimuli are always white, against a dark background.

In the interest of historical completeness, the original, nine subtest, 360 item version of Halsteads (Halstead and Settlage, 1943) is depicted in Table I-H.

Item*		Correct response
# 1	I	"1"
# 2	III	"3"
# 3	I	"1"
# 4	IV	"4"
# 5	II	"2"
# 6	IV	"4"
# 7	I	"1"
# 8	II	"2"

Figure I-A

Subtest I of Halstead Category Test

(Adapted from Simmel and Counts, 1957, pp. 19-20.)

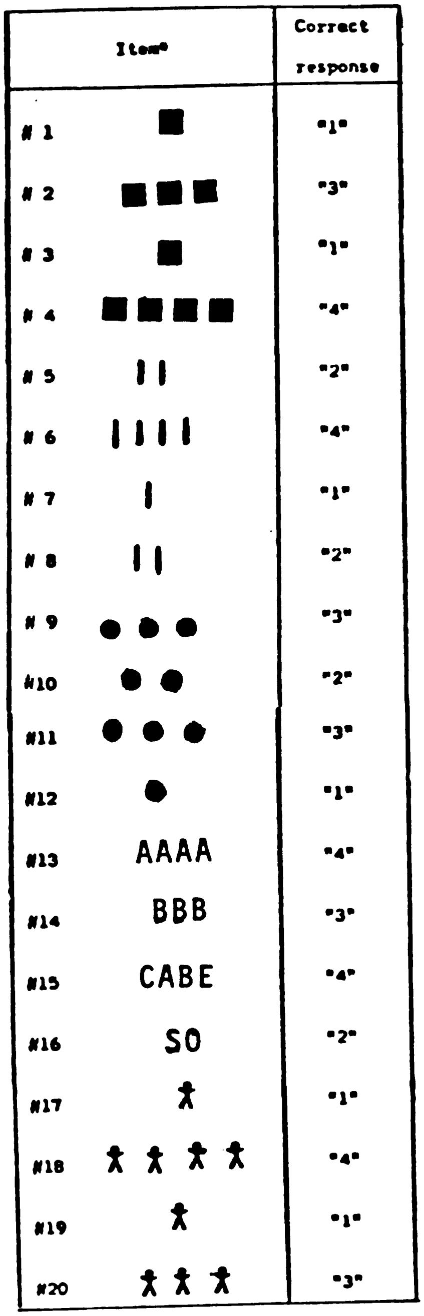


Figure I-B

Subtest II of Halstead Category Test

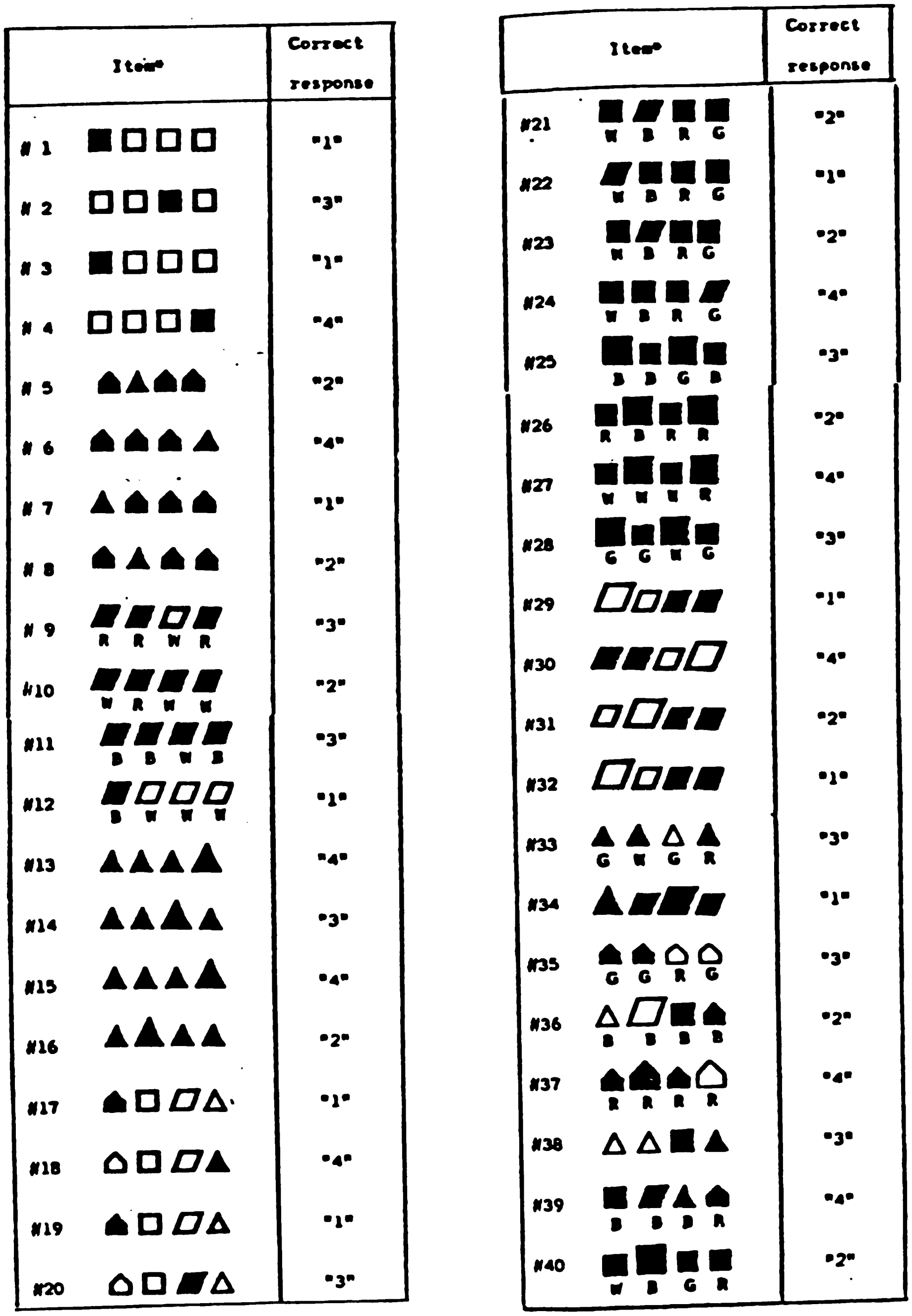
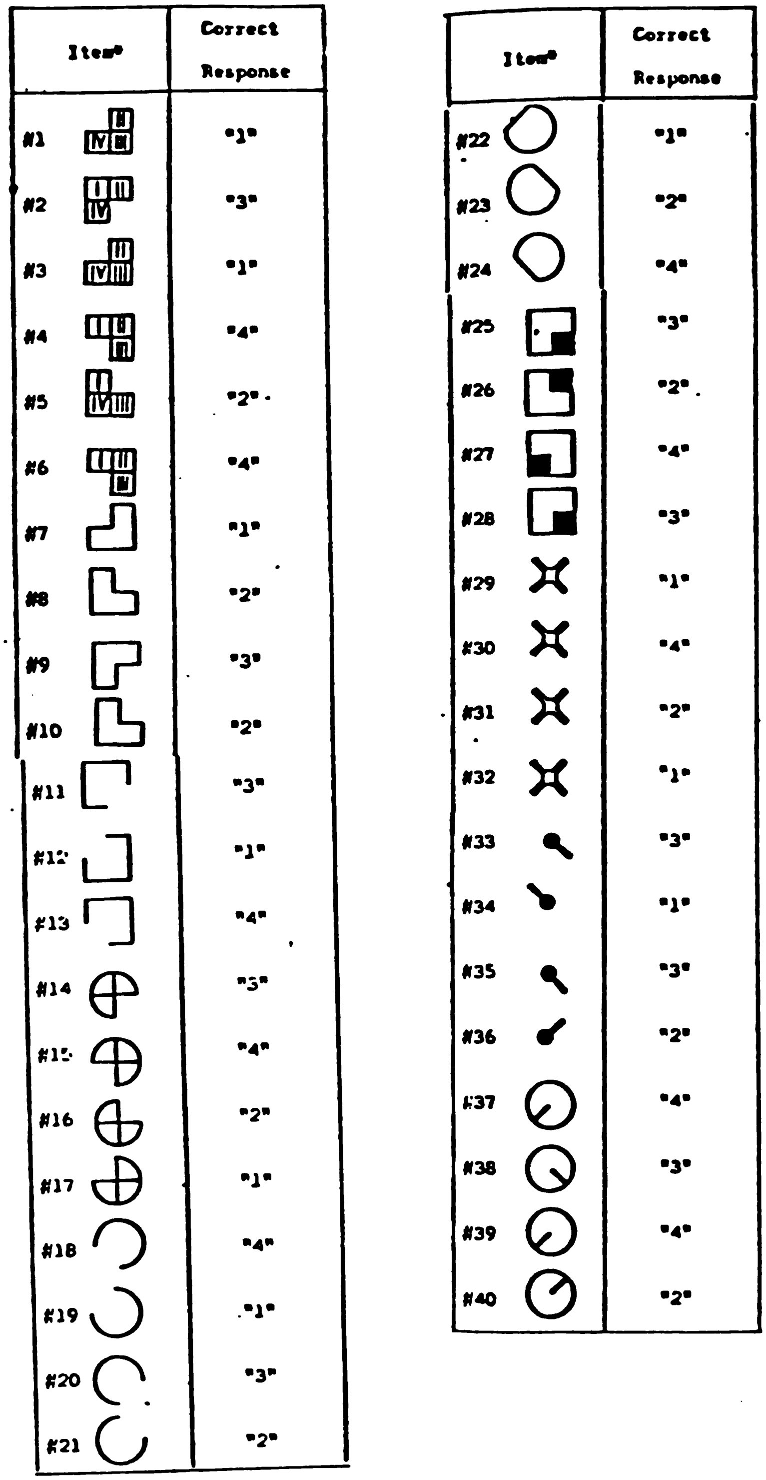


Figure I-C

Subtest III of Halstead Category Test

(Adapted from Simmel and Counts, 1957, pp. 28-29.)



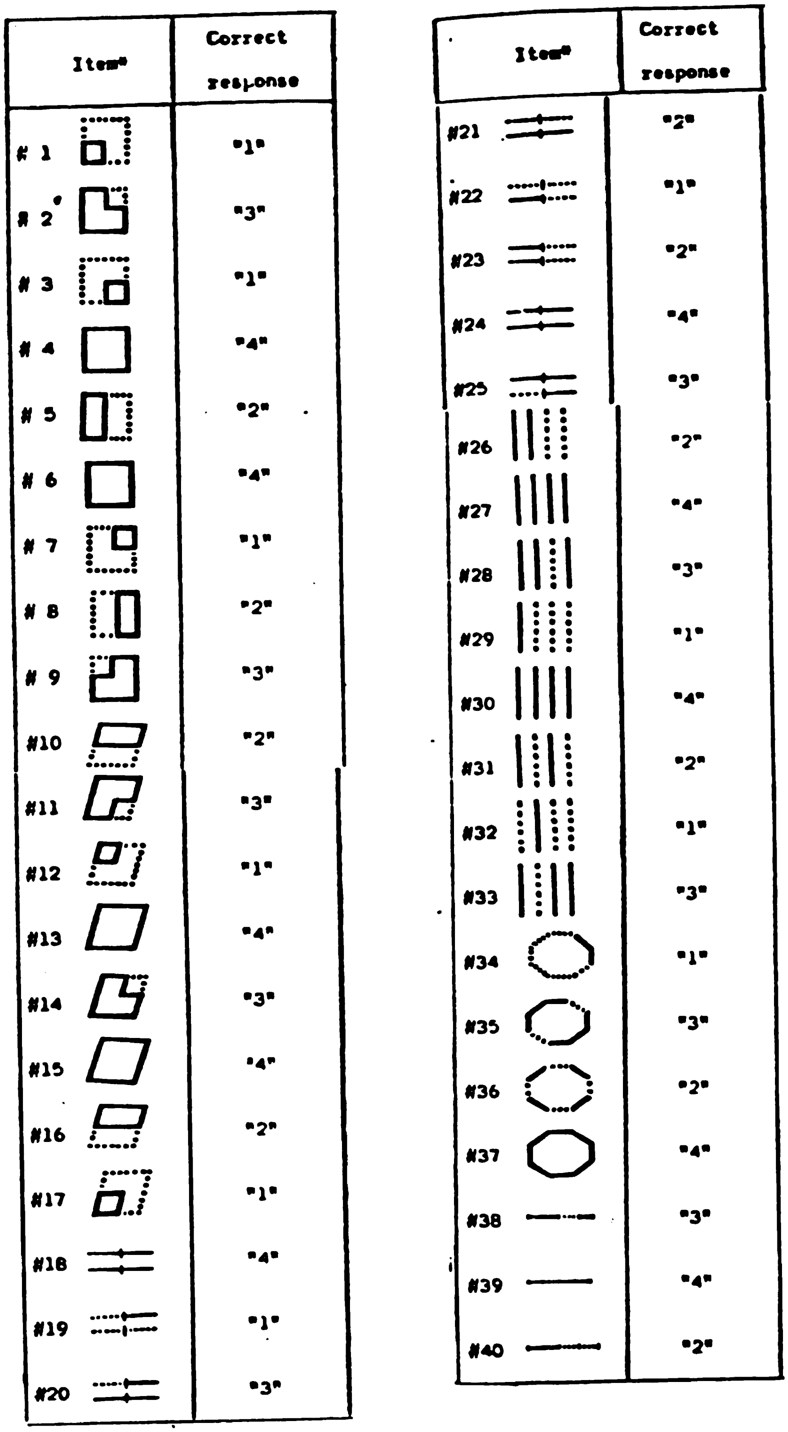
Item*	Correct Response
#1	"1"
#2	"3"
#3	"1"
#4	"4"
#5	"2"
#6	"4"
#7	"1"
#8	"2"
#9	"3"
#10	"2"
#11	"3"
#12	"1"
#13	"4"
#14	"3"
#15	"4"
#16	"2"
#17	"1"
#18	"4"
#19	"1"
#20	"3"
#21	"2"

Item*	Correct Response
#22	"1"
#23	"2"
#24	"4"
#25	"3"
#26	"2"
#27	"4"
#28	"3"
#29	"1"
#30	"4"
#31	"2"
#32	"1"
#33	"3"
#34	"1"
#35	"3"
#36	"2"
#37	"4"
#38	"3"
#39	"4"
#40	"2"

Figure I-D

Subtest IV of Halstead Category Test

(Adapted from Simmel and Counts, 1957, pp. 52-53.)



Item*	Correct response
# 1	"1"
# 2*	"3"
# 3	"1"
# 4	"4"
# 5	"2"
# 6	"4"
# 7	"1"
# 8	"2"
# 9	"3"
#10	"2"
#11	"3"
#12	"1"
#13	"4"
#14	"3"
#15	"4"
#16	"2"
#17	"1"
#18	"4"
#19	"1"
#20	"3"

Item*	Correct response
#21	"2"
#22	"1"
#23	"2"
#24	"4"
#25	"3"
#26	"2"
#27	"4"
#28	"3"
#29	"1"
#30	"4"
#31	"2"
#32	"1"
#33	"3"
#34	"1"
#35	"3"
#36	"2"
#37	"4"
#38	"3"
#39	"4"
#40	"2"

Figure I-E

Subtest V of Halstead Category Test

(Adapted from Simmel and Counts, 1957, pp. 74-75)

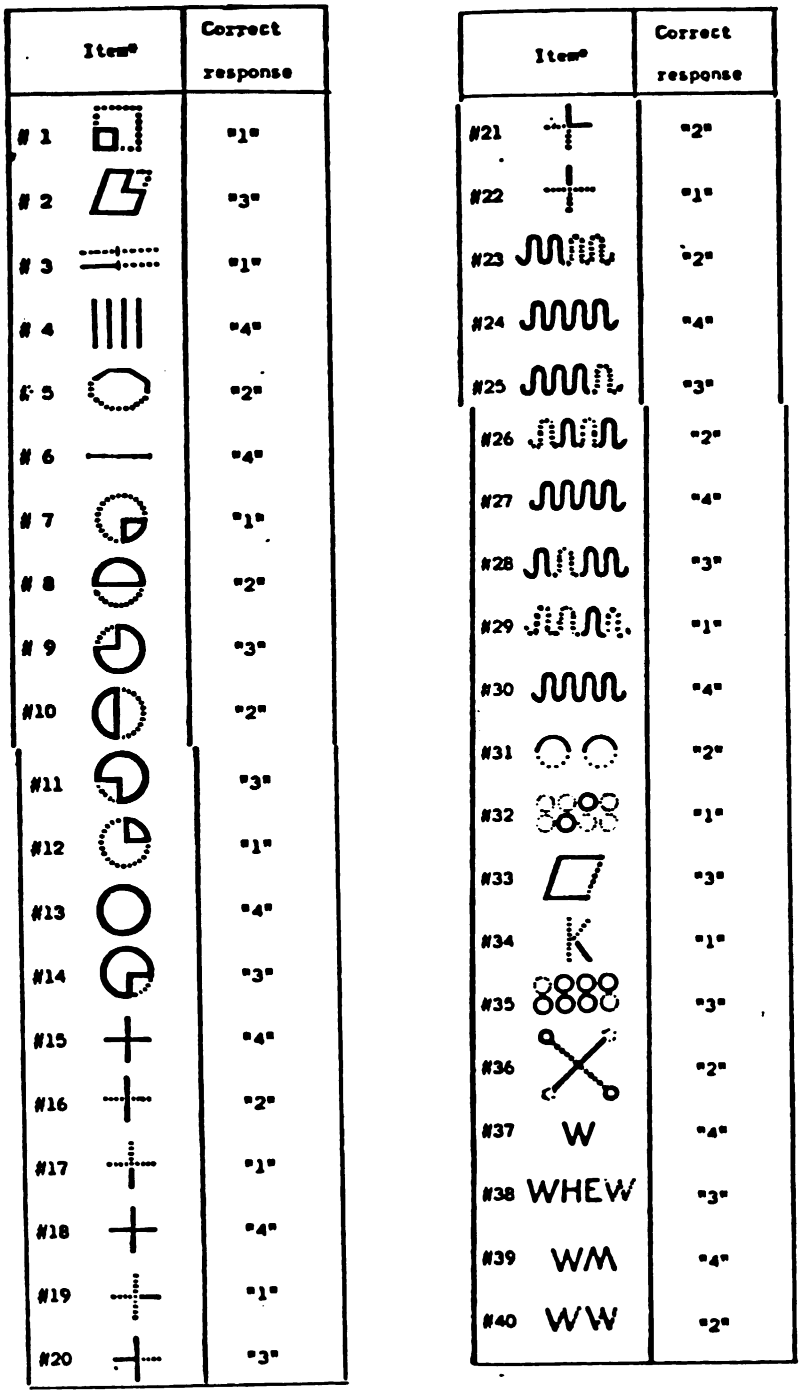


Figure I-F

Subtest VI of Halstead Category Test

(Adapted from Simmel and Counts, 1957, pp. 98-99.)

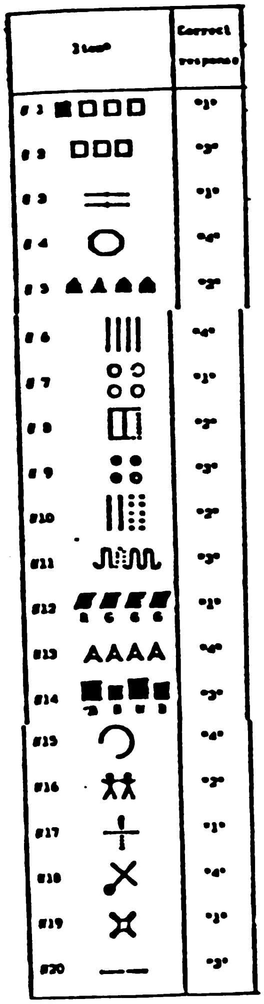


Figure I-G

Subtest VII of Halstead Category Test

(Adapted from Simmel and Counts, 1957, p. 119.)

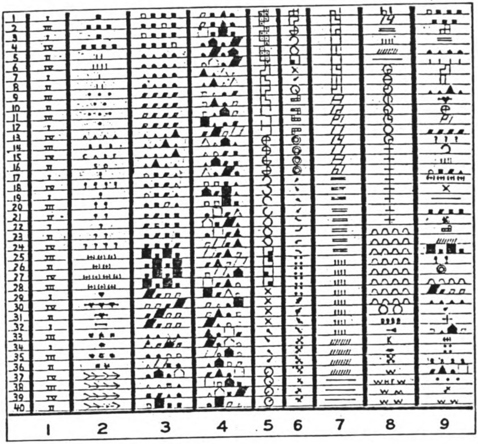


Figure I-H

Original Version of Halstead Category Test

(Adapted from Halstead and Settlage, 1943, p. 491.)

REFERENCES

REFERENCES

Adams, R. L., and Trenton, S. L. Development of paper-and-pen form of the Halstead Catagory Test. *Journal of Consulting and Clinical Psychology*, 1981, *49* (2), 298-299.

Anastasi, A. *Psychological testing*. Macmillan, 1976.

Barrett, E. T., Wheatley, R. D., and Laplant, R. J. A brief clinical neuropsychologic screening battery: statistical classification trials. *Journal of Clinical Psychology*, 1982, *38* (2), 375-377.

Battersby, W. S. Neuropsychology of higher processes: Cerebral damage and visual perception. In *Progress in clinical psychology*. New York: Grune and Stratton, 1956.

Beaumont, F. G. The validity of the Category Test administered by on-line computer. *Journal of Clinical Psychology*, 1975, *31*, 458-462.

Boyle, G. J. Shortened Halstead Category Test. *Australian Psychologist*, 1975, *10*, 81-85.

Brunswik, E. *Perception and the representative design of psychological experiments*. Berkeley: University of California Press, 1956.

Calsyn, D. A., O'Leary, M. R., and Chaney, E. R. Shortening theCategory Test. *Journal of Consulting and Clinical Psychology*, 1980, *48* (6), 788-789.

Campbell, D. T., and Stanley, J. C. *Experimental and quasi-experimental designs for research*. Chicago: Rand Mc Nally, 1963.

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

Cohen, J. A. A factor-analytically based rationalle for the Wechsler Adult Intelligence Scale. *Journal of Consulting Psychology*, 1957, *21*, 451-459.

Cohen, J., and Cohen, P. *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, New Jersey: Lawrence Erlbaum Associates 1975.

Cronbach, L. J. Coefficient alpha and the internal structure of tests. Psychometrika, 1951, 16 (3), 297-334.

Cronbach, L. J. Correlation between persons as a research tool. In O.H. Mowrer (Ed.), Psychotherapy, theory, and research. New York: Ronald, 1953.

Cronbach, L. J. Processes affecting scores on "understanding of others" and "assumed similarity." Psychological Bulletin, 1955, 52, 177-194.

Cronbach, L. J. Proposals leading to analytic treatment of social perception scores. In R. Taguiri and L. Petrullo (Eds.), Person-perception and interpersonal behavior. Stanford, CT: Stanford University Press, 1958.

Cronbach, L. J., and Glesner, G. Assessing similarity between profiles. Psychological Bulletin, 1953, 50, 456-474.

Doehring, D. G., and Reitan, R. M. Behavioral consequences of homonymous visual field defects in brain-injured adults. Journal of Corporative and Physiological Psychology, 1961, 54, 489-492.

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

Erickson, R. C., Calsyn, D. A., and Scheupbach, C. S. Abbreviating the Halstead-Reitan Neuropsychological Test Battery. Journal of Clinical Psychology, 1978, 34 (4), 922-926.

Filskov, S. B., and Goldstein, S. G. Diagnostic validity of the Halstead-Reitan Neuropsychological Battery. Journal of Consulting and Clinical Psychology, 1974, 42 (3), 382-388.

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

Golden, C. J. The identification of brain damage by an abbreviated form of the Halstead-Reitan Neuropsychological Battery. Journal of Clinical Psychology, 1976, 32 (4), 821-826.

Golden, C. J., and Anderson, S. M. Short form of the speech sounds perception test. Perceptual and Motor Skills, 1977, 45, 485-486.

Golden, C. J., Koperman, S. K., MacInnes, W. D., and Moses, J. A., Jr. Cross-validation of an abbreviated form of the Halstead Category Test. Journal of Consulting and Clinical Psychology, 1981, 49 (4), 606-607.

Goldstein, K. The significance of the frontal lobes for mental performance. Journal of Neurology and Psychopathology, 1936, 17, 27-40.

Goldstein, K. Clinical and theoretical aspects of lesions of the frontal lobes. Archives of Neurology and Psychiatry, 1939, 41, 865-867.

Goldstien, K. Human nature. Cambridge, Mass: Harvard University Press, 1940.

Goldstein, K. The two ways of adjustment of the organism to cerebral defects. Journal of the Mt. Sinai Hospital, 1942, 9, 504-513.

Goldstein, K. Mental changes due to frontal lobe damage. Journal of Psychology, 1944, 17, 187-208.

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

Gregory, R. J., Paul, J., and Morrison, M. W. A short form of the Category Test for adults. Journal of Clinical Psychology, 1979, 95, (4), 795-798.

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

Halstead, W. C. Preliminary analysis of grouping behavior in patients with cerebral injury by the method of equivalent and non-equivalent stimuli. American Journal of Psychiatry, 1940, 96 (2), 1263-1294.

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

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

Halstead, W. C. Specialization of behavioral functions and the frontal lobes. Research Publications: Association for Research in Nervous and Mental Disease (Vol. 27), 1948.

Halstead, W. C. Biological intelligence. Journal of Personality, 1951, 20, 118-130.

Halstead, W. C., Apter, N. S., & Heimburger, R. F. Impaired cerebral functions in essential hypertension. American Journal of Psychiatry, 1951, 107 (11).

Halstead, W. C., Carmichael, H. T., & Bucy, P. C. Prefrontal lobotomy: A preliminary appraisal of the behavioral results. American Journal of Psychiatry, 1946, 103, 217-228.

Halstead, W. C., and Chapman, L. F. Formation of auditory-visual pattern associations in humans with lesions of the temporal lobe. American Journal of Physiology, 1954, 179 (3).

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

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

Halstead, W. C., Symmes, D., & Chapman, L. F. The fusion of intermitten white noise. Journal of the Acoustical Society of America, 1955, 27 (3), 470-473.

Hammond, K. R. The psychology of Egon Brunswik. New York: Holt, Rinehart, and Winston, 1966.

Hanfmann, E., & Kasanin, J. A method for the study of concept formulation. Journal of Psychology, 1937, 3, 521-540.

Hartmann, H. Ego psychology and the problems of adaptation. New York: International Universities Press, 1958.

Heaton, R. K., & Crowley, T. J. Effects of psychiatric disorders and their somatic treatments on neuropsychological test results. In Filskov, S.B., & Bull, T. J. (Eds.). Handbook of clinical neuropsychology. New York: John Wiley and Sons, 1981.

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

Kilpatrick, D. G. The Halstead Category Test of brain dysfunction: Feasibility of a short form. Perceptual and Motor Skills, 1970, 30, 577-578.

Kilpatrick, D. L., and Spreen, O. A revision of the Halstead Category Test for children aged 9 to 15. Psychology in the Schools, 1973, 10, 101-106.

Kimura, S. D. A card form of the Reitan-modified Halstead Category Test. _Journal of Consulting and Clinical Psychology_, 1981, _49_ (1), 145-146.

Kluver, H. Relational thinking in monkeys. _Psychological Bulletin_, 1929, _26_. 168-169.

Kluver, H. The equivalence of stimuli in the behavior of monkeys. _Journal of Genetic Psychology_, 1931, _39_, 3-27.

Kluver, H. The study of personality and the method of equivalent and non-equivalent stimuli. _Character and Personality_, 1936, _5_, 91-112.

Kohler, W. _The mentality of apes_. Translated by E. Winter. New York: Harcourt, Brace, and Company, 1925.

Lacks, P. B., Colbert, J., Harrow, M., and Levine, J. Further evidence concerning the diagnostic accuracy of the Halstead organic test battery. _Journal of Clinical Psychology_, 1970, _26_, 480-481.

Landsell, H., and Donnelly, E. F. Factor analysis of the Wechsler Adult Intelligence Scale subtests and the Halstead-Reintan Category and tapping tests. _Journal of Consulting and Clinical Psychology_, 1977, _45_ (3), 412-416.

Lenzer, I. I. Halstead-Reintan Test Battery: A problem of differential diagnosis. _Perceptual and Motor Skills_, 1980, _50_, 611-630.

Lewis, R. F., Nelson, R. W., and Eggersten, C. Neuropsychological test performances of paranoid schizophrenic and brain-damaged patients. _Journal of Clinical Psychology_, 1979, _35_, 54-59.

Lin, Y., and Rennick, P. M. Correlations between performance on the Category Test and the Wechsler Adult Intelligence Scale in an epileptic sample. _Journal of Clinical Psychology_, 1974, _30_, 62-65.

Luria, A. P. _The working brain_. New York: Penguin Press, 1973.

Luria, A. R., and Majorski, L. Basic approaches used in American and Soviet clinical neuropsychology. _American Psychologist_, 1977, _32_, 959-968.

Mack, J. C., and Carlson, N. J. Conceptual deficits and aging: The Category test. _Perceptual and Motor Skills_, 1978, _46_, 123-128.

Magnusson, D. _Test theory_. Reading, MA: Addison-Wesley, 1967.

Malec, J. Neuropsychological assessment of schizophrenia _versus_ brain damage: A review. _Journal of Nervous and Mental Disease_, 1978, _166_ (F), 507-516.

Matarazzo, J. D., Matarazzo, R. G., Wiens, A. N., Gallo, A. E., and Klonoff, H. Retest reliability of the Halstead Impairment Index in a normal, a schizophrenic, and two samples of organic patients. _Journal of Clinical Psychology_, 1976, _32_ (2), 338-349.

McCampbell, E., and DeFillippis, N. A. The development of a booklet form of the Category Test: A preliminary report. _Clinical Neuropsychology_, 1979, _1_ (2), 33-35.

Mc Fie, J., and Piercy, M. F. Intellectual impairment with localized cerebral lesions. _Brain_, 1952a, _75_, 292-311.

Mc Fie, J., and Piercy, M. F. The relation of laterality of lesions to performance on Weigl's sorting task. "_Journal of Mental Science_, 1952b, _98_, 299-305.

Meehl, P. E., and Rosen, A. Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. _Psychological Bulletin_, 1955, _52_, 194-216.

Mezzich, J. E., and Moses, J. A., Jr. Efficient screening for brain dysfunction. _Biological Psychiatry_, 1980, _15_ (2), 333-337.

Milner, B. Effects of different brain lesions on card sorting. _A.M.A. Archives of Neurology_, 1963, _9_, 90-100.

Prigatano, G. P., and Parsons, O. A. Relationship of age and education to Halstead test performance in different patient populations. _Journal of Consulting and Clinical Psychology_, 1976, _44_ (4), 527-533.

Rachlin, H. _Behavior and learning_. San Francisco: W.H. Freeman and Company, 1976.

Reed, H. B. C., Jr. and Reitan, R. M. A comparison of the effects of the ormal aging process with the effects of organic brain-damage on adaptive abilities. _Journal of Gerontology_, 1963a, _18_, 177-179.

Reed, H. B. C., Jr., and Reitan, R. M. Changes in psychological test performance associated with the normal aging process. _Journal of Gerontology_, 1963b, _18_, 271-274.

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

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

Reitan, R. M. The relationshhip of Halstead Impairment Index and the Wechsler-Bellevue total weighted score to chronologic age. (Paper presented at the ninth annual scientific meeting of the Gerontological Society, Inc., November 8, 1956a).

Reitan, R. M. Investigation of relationships between "psychometric" and "biologic" intelligence. *Journal of Nervous and Mental Disease"*, 1956b, *123*, 536-541.

Reitan, R. M. Differential reaction of various psychological tests to age. *Fourth Congress of the International Association of Gerontology*, 1957, *4*, 158-165.

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

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

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

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

Reitan, R. M. 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. I). New York: Academic Press, 1966, 153-218.

Reitan, R. M. Manual for administration of neuropsychological test batteries for adults and children. Indianapolis: Auther, 1969.

Reitan, R. M., & Davison, L. A. (Eds.). *Clinical neuropsychology: Current status and applications*. New York: John Wiley & Sons, 1974.

Rourke, B. P. Central processing deficiencies in children: Toward a developmental neuropsychological mode. Journal of Clinical Psychology, 1982, 4 (1), 1-18.

Russell, E. W. A multiple scoring method for the assessment of complex memory functions. Journal of Consulting and Clinical Psychology, 1975, 43 (6), 800-809.

Russell, E. W., Neuringer, C., and Goldstein, G. Assessment of brain damage: A neuropsychological key approach. New York: John Wiley, 1970.

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

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

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

Thurstone, L. L. Multiple-factor analysis. Chicago: University of Chicago Press, 1947.

Tucker, D. M. Lateral brain function, emotion, and conceptualization. Psychological Bulletin, 1981, 89 (1), 19-46.

Vega, A., Jr., and Parsons, O. A. Cross-validation of the Halstead-Reitan tests for brain damage. Journal of Consulting Psychology, 1967, 31 (6), 619-625.

Walsh, K. W. Neuropsychology: A clinical approach. New York: Churchill Livingstone, 1978.

Watson, C. G., Thomas, R. W., Andersen, D., and Felling, J. Differentiation of organics from schizophrenics at two chronicity levels by use of the Reitan-Halstead organic test battery. Journal of Consulting and Clinical Psychology, 1968, 32 (6), 679-684.

Watson, C., Thomas, R., Felling, J., and Anderson D. Differentiation of organics from schizophrenics with Reitan's sensory-perceptual disturbances test. Perceptual and Motor Skills, 1968, 26, 1191-1198.

Wechsler, D. The measurement of adult intelligence. Baltimore: Williams & Wilkins, 1944.

Wechsler, D. A standardized memory scale for clinical use. _Journal of Psychology_, 1945, _19_, 87-95.

Wechsler, D. _Manual for the Wechsler Adult Intelligence Scale_. New York: Psychological Corporation, 1955.

Weigl, E. On the psychology of the so-called processes of abstraction. _Journal of Abnormal and Social Psychology_, 1941, _36_, 1-33.

Wheeler, L. W., Burke, C. J., and Reiten, R. M. An application of discriminant functions to the problem of predicting brain damage using behavioral variables. _Perceptual and Motor Skills_, 1963, _16_, 417-440.

Winer, B. J. _Statistical principles in experimental design_. New York: McGraw-Hill, 1962.

Wood, W. W., and Strider, M. A. Comparison of two methods of administering the Halstead Catagory Test. _Journal of Clinical Psychology_, 1980, _36_ (2), 476-479.