THE MEMORY SPAN PARADIGM: ITS USE FOR THE ANALYSIS OF MENTAL RETARDATION

Thesis for the Degree of Ph. D.

MICHIGAN STATE UNIVERSITY

BRUCE LOWELL BACHELDER

1970

This is to certify that the

thesis entitled

The Memory Span Paradigm: Its Use for the Analysis

of Mental Retardation

presented by

Bruce Lowell Bachelder

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Psychology

M. Ray Denny

Major professor

Date Nov. 19, 1970

O-7639

ABSTRACT

THE MEMORY SPAN PARADIGM: ITS USE FOR THE ANALYSIS OF MENTAL RETARDATION

By

Bruce Lowell Bachelder

Two experiments were done, each of which used the sequential auditory memory span paradigm. The goal of Experiment 1 was to explain the materials effect, that is, the fact that digit spans are usually larger than common-word spans. It was hypothesized that digit spans are larger than word spans because the average digit response is larger than the average word response (the response size hypothesis).

The Ss were 30 retarded and borderline adults living in a home for the retarded ($\overline{CA}$=25.1, $\overline{IQ}$=69.83). Each S attempted to reproduce a constant number of digit and common-word strings ranging in size from 1-7 verbal units. The dependent variable was the number of correctly produced response strings.

Evidence of the presence of large responses was sought in the digit and word response-strings; the results were consistent with expectations. Digit transition numbers (the number of string sizes which S produced both correctly and incorrectly) were larger than word transition numbers ($p<.05$). The intrastring error rate (the rate of errors in the second and penultimate positions of incorrect response strings which have a correct first or second verbal unit, respectively) was lower for digit strings than for word strings ($p<.01$). It was concluded that the response size hypothesis is an adequate explanation of the materials effect.

Practice within the experimental session produced a steady decline in performance ($p<.01$). Practice did not interact with IQ, age, memory span, or size of materials effect (digit performance minus word performance). The effect of practice was attributed to boredom or response habituation.

The goal of Experiment 2 was to test the sufficiency principle which states that two constructs are sufficient to account for individual differences in the memory span paradigm: the S-R repertoire and the size of memory span. According to the sufficiency principle, experimental variables do not interact with levels of memory span; the results were essentially as predicted.

The Ss were 84 adult residents of a home for the retarded ($\overline{CA}$=27.5, $\overline{IQ}$=71.1). Ss were divided into a high IQ and a low IQ group ($\overline{IQ}$s=82.7 and 59.6, respectively). Digit strings were presented in a Rate X Grouping factorial design (½, 1, 2, and 4 digits per second X single vs. grouped, that is, pairwise presentation). Each S served in one treatment combination. Each S was classified according to size of digit span, and each treatment combination contained one S from each of six levels of memory span. The dependent variable was S's digit span as measured by the staircase method (adapted from the psychophysical technique).

As hypothesized, slow rates facilitated the performance of high IQ Ss ($p<.01$) but had no effect on low IQ Ss; the Rate X IQ interaction was significant ($p<.01$). Grouping facilitated the performance of all Ss ($p<.01$).

Within the high IQ Ss, neither rate nor grouping interacted with memory span level. Among low IQ Ss, memory span and grouping

did not interact; but, contrary to prediction, the Rate X Memory Span interaction was significant ($p<.05$). This latter result appeared to be due to error of classification of memory span, as 39 out of 42 Ss actually performed just as predicted. Performance declined as a function of practice ($p<.01$), and practice did not interact with size of memory span. It was concluded that the sufficiency principle was valid for the present experiment.

A concept of elicitation span (similar to memory span but more general) was advanced as an elaboration of the traditional S-R model of behavior. The present experiments were organized and discussed in terms of the elaborated S-R model; and it was suggested that the elaborated S-R model might be useful as a model of intelligence, language, and retardation.

THE MEMORY SPAN PARADIGM: ITS USE FOR THE ANALYSIS OF MENTAL RETARDATION

By

Bruce Lowell Bachelder

A THESIS

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

DOCTOR OF PHILOSOPHY

Department of Psychology

1970

ACKNOWLEDGMENTS

I would like to express my appreciation to Mr. Andreas Heath of Caro State Home and Training School for his original permission to work with the residents of the Home. I also thank the Home's Superintendent, Dr. Joseph C. Denniston, who granted me permission to work a part time schedule which allowed me time to collect data and write the thesis.

Dr. Marjorie Clos of Caro State Home and Training School provided special encouragement through the difficult writing phase. I thank the cottage personnel whose assistance in scheduling residents made data collection efficient. My dissertation committee are especially appreciated for the freedom they allowed me in developing this thesis.

Finally, I owe considerable debt to my wife for her comments on style, her typing, and her tolerance.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

In the memory span experiment, E presents a string of words or digits; and S attempts to repeat them in the same order presented by E. The average college student can repeat strings of one to six digits quite easily but begins to fail consistently on strings of seven or more digits. The largest string which S can repeat reliably is called his memory span. The size of memory span is directly related to IQ and mental age, and a digit span of less than 5 or 6 is diagnostic of brain damage in otherwise normal individuals (Wechsler, 1958).

The modern experimental use of variations of the memory span paradigm has been for testing theories of memory in normal and retarded subjects. The memory span phenomenon has been interpreted in terms of a decay theory of memory such as that of Brown (1958) which holds that memory traces decay rapidly with the passage of time (seconds). Brown theorized that the memory span is limited because the memory traces of early stimuli decay during presentation of subsequent stimuli. The fact that memory span is related to intelligence has been explained in terms of the strength and duration of the memory trace. Ellis (1963) and Scott & Scott (1968) have hypothesized that the memory traces of retarded Ss are weaker and decay more rapidly than the memory traces of normal Ss.

This paper presents an alternative interpretation and model

of the memory span phenomenon. The new interpretation places the memory span phenomenon within the stimulus-response tradition of behavior analysis. The term elicitation span (ES) is substituted for memory span because the span phenomenon is not explained in terms of memory decay.

The Model

The model is a simple elaboration of the traditional stimulus-response (S-R) conception of behavior.[1] In its least complex form, the ES paradigm consists of E presenting one word and S producing one word. The paradigm is made more complex by increasing the number of words E presents and which S must reproduce. In its simplest form, the interaction between E and S is readily symbolized by the usual S-R model of behavior; i.e., E presents a stimulus which elicits a response from S. The more complex forms of the paradigm imply elaboration of the S-R model to $S_1S_2S_3 \ldots S_n$--$R_1R_2R_3 \ldots R_n$. According to this elaborated S-R model ($S\text{-}R_{elaborated}$, or $S\text{-}R_e$), behavior is produced, not as simple S-R events, but as complex sequences of responses (response strings) in response to a stimulus complex. The response string is not to be conceived as a chain of associated responses. There is no implication of association between the R_ns in the response string. The sequential arrangement can be, and frequently is,

---

[1] This statement must be qualified to the extent that the present conceptions of S and R are similar to those of Denny (1966, 1970 a,b). According to Denny, stimulus and response are abstract concepts which are inferred from behavior and objects or events. In general, the distinction between Denny's S and R and the traditional S and R is not important in this paper. At certain times, however, responses (for example, large responses) are inferred from behavior and stimulus events; this practice departs from strict stimulus-response analysis.

completely novel because of the novelty of the stimulus complex.

The literature suggests the following principles: (a) The ES grows during the developmental period. (b) Individuals differ in size of ES and these differences are related to intelligence. (c) The S-R_e model applies to a wide variety of behaviors. (d) Practice affects only the nature of the S-R unit, it does not increase the size of ES.

An additional principle has no support from the available literature but is the topic of Experiment 2 below: (e) To account completely for individual differences in the elicitation span paradigm, it is necessary to specify only the size of ES and the content and character of S's repertoire of S-R units. This principle is called the sufficiency principle.

The development of ES. It is assumed that ES grows during childhood and reaches a plateau during adolescence. This assumption is based on the fact that digit span and other memory span scores have been shown repeatedly to follow this developmental sequence.

Wechsler (1958) studied performance on the digit span subtests of his IQ tests, from which he concluded that the size of digit span increases during childhood and reaches a plateau in adolescence or young adulthood. Wechsler also found small increases in digit span to age 25 and a decline with increasing age which was greater than the same decline of the other subtests.

Jacobs (1887), in one of the earliest studies of the memory span, tested digits, letters, and nonsense syllables. In each case, Jacobs found a positive relation between age and size of

span which appeared to reach an asymptote at 18-20 years.

Korst & Irwin (1968 a,b) measured memory span by means of a standard digit test and by S's immediate memory for two short prose passages. The two tests were pooled for a total memory span score. Korst & Irwin found that span scores are positively related to age in school-age children (grades four through seven) and in retarded children ages 6 through 16.

Gundlach, Rothschild, & Young (1927) used an unusual test of memory span. The stimuli were a circle of 10 lights. A stimulus string was a random sequence of light flashes around the ring, and S responded by touching the lights in the order they flashed. Gundlach et al. found that performance in this task is a function of age and appears to reach a plateau by about 14-16 years.

Lumley & Calhoon (1934) measured the sequential auditory memory span for common one-syllable words in 946 school children and found that span scores are directly related to age (grade placement). Each score below is the average score over several presentation rates (rate had no important effect):

third grade - 2.1
fifth grade - 3.2
seventh grade - 3.7
ninth grade - 4.1
twelfth grade - 4.7

Individual differences in ES. It is assumed that the size of ES is a stable characteristic of the individual and is positively related to intelligence. This principle is suggested by the extensive use of the span test as a clinical tool to detect mental retardation (Wechsler, 1958), brain damage (Hunt, 1943), aphasia (Eisenson, 1954), psychosis (Wechsler, 1958), and to measure

psycholinguistic abilities (Kirk, McCarthy, & Kirk, 1968).

Wechsler (1958) stated: "Except in the cases of special defects or organic disease, adults who cannot retain 5 digits forward and 3 backward will be found, in 9 cases out of 10, to be feebleminded or mentally disturbed [p. 71]." Galton (1887) reported that retarded persons have smaller than normal digit spans.

Pilot work for the present investigation clearly indicated that persons living in a residential facility for the mentally retarded vary along a continuum of word spans ranging from .5 to 5 units. This continuum corresponds roughly to the range between profoundly retarded people and people of borderline normal intelligence.

Butterfield (1968) matched normal and retarded Ss on mental age and sex and found no difference in raw score or variability on the WISC digit span subtest. Since the retarded Ss were older than the normal Ss, the retarded Ss were deficient in digit span compared with normal Ss of the same ages.

Wechsler (1958) stated that as a measure of intelligence, the digit span subtest is among the poorest. Jensen (1970) pointed out, however, that Wechsler's digit span test is short and its reliability is among the lowest of all Wechsler's subtests. According to Jensen, the correlation between digit span and the WAIS IQ is approximately .75 when corrected for attenuation; and on the WAIS tests the digit span subtest has a factor loading of about .80 on the "g" factor which Wechsler (1958) extracted in a factor analysis. Jensen further reported that lengthening the digit span test produces digit span reliabilities of .90-.96.

Brener (1940) measured the memory spans of 40 college

students for a variety of verbal materials including digits, abstract words, concrete words, colors, sentences, and letters. His procedure was the sequential presentation of stimuli at a rate of one stimulus each 2 seconds. Brener used both visual and auditory presentation both of which produced very similar spans (see Table 1, Experiment 1).

Brener's results provide another type of evidence which suggests that the ES is constant and general for an individual and differs in size from one person to the next. He reasoned:

> If there is only one general memory span factor involved, that is, if the individual's relative position in the group depends entirely on one ability that is called for in all of the tests and is not influenced by the type of material on which the span is tested, we would expect the distribution of the mean standard scores for all subjects to be fairly well scattered from high positive to high negative values. Likewise, we would expect any subject to have approximately the same standard score on all tests, with little variability in the score [p. 475].

Brener's 40 Ss produced mean standard scores (averaged across materials) which ranged from -1.30σ to $+1.77\sigma$; that is, Ss differed widely in average span scores. Brener found also that each S's standard score is quite stable across materials; the mean standard error of the standard scores (measured by different materials) was .091. The small average standard error means that each S scored at his characteristic span level regardless of the testing materials.

The generality of the S-R_e model. The S-R_e model implies that evidence of the span phenomenon should be found for a variety of S-R units and that estimates of ES should be quite similar for different materials. Some relevant data were reviewed earlier in another context: Jacobs (1887), for example, found similar

developmental growth curves for digits, letters, and nonsense syllables. Gundlach et al. (1927) found span limits and the developmental growth sequence for the circle of lights task; Lumley & Calhoun (1934), for common words. Brener (1940) approached the generality problem directly.

Brener reasoned that if there were an underlying "memory span," his various tests would be highly correlated. He factor analyzed his data and extracted a general "memory span" factor on which all subtests had significant loadings.

The usual span test uses stimuli and responses which are topographically similar; that is, S imitates E's stimuli rather exactly. Brener used more general tests, however, in which the stimuli and S's responses were not topographically similar. The mean span scores were quite similar (5.24-7.98) for a variety of S-R relations:

Stimuli	Responses
(a) Spoken words (letters and digits; concrete and abstract words)	spoken words
(b) Visually presented letters	spoken letter names
(c) Visually presented color samples	spoken color names

There is also evidence that the same ES constant applies to other less verbal behavioral paradigms. Brener's subtest, called the "memory for commissions," consisted of visual presentation of stimulus cards, each of which had simple instructions printed on it. Each card required that S perform a response with three aspects ("operations") as illustrated by this sentence used by Brener: "Put a comma under the A." The three aspects are (a) put a comma, (b) under, and (c) the A. The S responded by carrying

out the instructions on a card which had two letters printed on it. There are two important characteristics of the task: The stimuli are not topographically similar to the responses and the responses are nonverbal. Brener found that the number of cards correctly responded to was approximately 2.42. Since each card symbolized three operations which S had to produce, and assuming that each operation was a response, we can estimate the average elicitation span as 3 x 2.42=7.26. This figure is very close to the figure obtained for digits and consonants: 7.98 and 7.30, respectively.

Brener's data support the general span concept, but at the same time offer some difficulty for the S-R_e model. Even though spans for different materials are highly related, they are consistently different. Digits produce higher span scores than common words, a fact which must be reconciled with the assumption of a constant ES for the individual. This problem is the topic of Experiment 1 below.

The effect of practice. ES is assumed to be constant for the individual (with the exception of developmental changes). This assumption is inconsistent with a practice effect. Blankenship (1938) reviewed the literature on memory span and reported that the data relevant to a practice effect are inconclusive. Scott & Scott (1968) reviewed the literature on short-term memory in the retarded and concluded that practice causes no improvement in memory span in the sense of improving an underlying ability.

There are some recent data on practice which are relevant but which do not involve measures of the absolute span which is the topic of this paper. Most studies of short-term memory

present supraspan strings and measure S's performance in terms of errors at specific positions within a response string. Two studies of this type suggest that practice effects are restricted to changes in S-R units and not to changes in ES.

Melton (1963) reported a study which was a modified replication of an earlier study by Hebb (1961). In each study, E presented large numbers of stimulus strings just beyond span. Most of the strings occurred just once, but some of the strings occurred repeatedly at varying intervals. The result was that the frequency of perfect string productions increased with practice, but only for the repeated strings. The frequency of correct productions increased as a direct function of the number of repetitions of a specific string. Melton's interpretation of the effect was that, through repetition of specific sequences, S formed larger "chunks" (Miller, 1956) from the individual response items (the chunk is operationally similar to the present response). In other words, S formed associated pairs or triplets or even larger groups of response units and thereby became able to reproduce greater numbers of supraspan strings. The important finding is that performance did not improve on randomly generated strings; facilitation was restricted to specific and repeated strings. These data are consistent with the assumption that practice affects only the nature of the S-R unit and not the size of ES conceived as an underlying general ability.

In summary, it appears that the S-R_e model applies to a wide variety of behaviors which are not necessarily verbal (for example, the test for commissions and the touching of lights). The work on

the relation between intelligence and span size and Brener's extensive correlational and factor analysis of spans for many different classes of S-R units support the concept of the ES as an underlying ability. On the other hand, practice should have no effect on size of ES; but this remains to be shown convincingly. Of course, proving that there is no improvement with practice will always be difficult because it amounts to proving the null hypothesis. The actual "test" of the principle will be the continuing success of the theorist to attribute any improvements which occur with practice to changes in S-R units rather than to changes in the size of ES. Finally, the fact that digits consistently produce higher scores than other words (the materials effect) must be resolved if we are to retain the concept of the constant-valued underlying ES.

The Study of Mental Retardation

The overriding concern of this paper is to outline a new approach to the study of mental retardation. Since the size of span is related to intelligence, analysis of the span phenomenon, especially with respect to individual differences in span, should be relevant to an understanding of retardation. Three of the most typical approaches to the experimental analysis of mental retardation are discussed below. Each approach, as well as the $S\text{-}R_e$ approach, is directly related to the goals of the individual researcher. Each approach is valid in its own way.

The search for behavioral deficits. The search for behavioral deficits is called the comparative approach. Research designs compare normal and retarded Ss matched on either CA (Ellis, 1969,

1970), MA (Zigler, 1969), or use both types of comparison groups (Denny, 1964). The goal of the comparative approach is to point up differences between retarded and normal behavior under specified conditions and to seek empirical support for hypothetical deficits such as deficits of attention (Zeaman, 1963), short-term memory (Ellis, 1970), and stimulus organization ability (Spitz, 1966).

The comparative approach can be conceived as the first phase of a two-phase analysis of mental retardation. In the first phase, the goal of research is to specify the exact deficit(s) of retardates; in the second phase, to subject the deficit(s) found in the first phase to further analysis and to specify the empirical and theoretical relations among the deficit(s) and intelligence and basic behavior theory.

The present approach, based upon the S-R_e model, is an example of the second phase of the comparative approach. It is assumed that a basic deficit in digit and word span has been established for some or all retardates. The goal of research, then, is to establish the empirical and theoretical relation of this span deficit to intelligence, retardation, and basic behavior theory.

Functional analysis of retarded development. The functional analysis approach has been described by Bijou (1966): "From this point of view a retarded individual is one who has a limited repertory of behavior shaped by events that constitute his history [p. 2]." Research based upon this conception of retardation seeks to demonstrate that the systematic application of positive

reinforcers in a carefully structured training program produces improvements in behavior, that is, makes S's behavior more "normal" (see, for example, Bricker, 1970; Sidman & Stoddard, 1966; Spradlin & Girardeau, 1966).

The functional analysis approach seeks to demonstrate the sufficiency of a simple conception of retardation: Retardates, for a variety of reasons, have not had full benefit of a normal reinforcement history; and proper training remediates behavior deficits. In this sense, the functional analysis resembles the S-R_e approach which attempts to establish the sufficiency of the ES concept to account for individual differences in intelligent behavior.

The functional analysis approach does not have a capacity variable; Ss are conceived to differ primarily in terms of behavioral (S-R) repertoire rather than in their capacity to develop a repertoire. The S-R_e approach has a capacity variable (ES) which presumably limits training in certain theoretically specific ways. It is agreed that the reinforcement history is an important developmental variable, but the S-R_e model postulates that individuals differ in their capacity to take advantage of a normal environment.

The study of basic processes in retardates. Zeaman (1967) and Scott & Scott (1968) have argued that the comparative approach is a very inefficient and potentially misleading way to study the behavior of retardates. The comparative approach is misleading because the intelligence factor is not an experimental factor; we really do not know why normal and retarded Ss differ when they do differ. A behavior deficit in retarded Ss may be due to a deficit in a basic behavior process or to a deficit in experience. The

alternative that Zeaman and Scott & Scott propose is the study of retarded Ss in their own right. Both Zeaman and Scott & Scott argue that attempts to improve the lot of retardates must rely on principles of retarded behavior and need have no relation to the behavior of normal Ss. Scott & Scott also point out that an understanding of behavior processes in normal and retarded Ss would provide a basis of interpretation of deficits found in comparative studies.

It is a valid argument that we can effectively study retarded behavior as distinct from normal behavior, but this very argument implies a dichotomization of subjects. Retarded and normal Ss seem to exist on a continuum of intelligence which is rather Gaussian in form, except for IQs <50 (Jensen, 1970; Robinson & Robinson, 1965). The S-R_e approach encompasses a continuum of individual differences in ES. Behavior differences are conceived as continuous and quantifiable. No dichotomization of Ss is implied.

The S-R_e approach. These three aforementioned approaches are being used with considerable success, but they all suffer from one or more of the following problems: (a) They have no concept of ability which can be stated in basic theoretical terms. (b) They treat the continuum of intelligence as if all persons could be dichotomized into the normal and the retarded. (c) They conceive of retardation as a unitary, rather than complex, phenomenon.

The S-R_e model responds, at least in part, to each of these problems: (a) It presumes a theoretical concept of ability, the ES, which exists independent of specific S-R repertoires. (b) It

treats Ss as lying on a continuum of ES ability. (c) It does not use the concept of the ES to define retardation; the concept of the ES is used to define Ss for the purpose of examining individual differences in ES. Because of the rather close relation between ES and intelligence and retardation, the S-R_e approach does, of course, study retarded Ss; but it also allows for the possibility of other types of retardation such as deficiencies in early experience and deficits of attention and associative ability. It is clearly acknowledged, however, that one of the goals of a research program based on the S-R_e model should be to establish the sufficiency of the ES as a conception of individual differences in intelligence. Against this theoretical base, any other deficits should be amenable to clear specification. The S-R_e approach is not comparative; it does not necessarily expect to find differences between "retarded" and "normal" Ss. In fact, by presenting tasks which are matched to S's elicitation span it may be possible to eliminate differences between Ss, or in other words, to show that the essential difference between Ss is the size of ES.

Experiment 1:

The Effects of Practice and Stimulus Class on the Memory Spans of Borderline and Retarded Ss

According to the S-R_e model, the size of ES is a constant and general limit on the ability of a given S to form novel sequences of discrete responses. The best evidence for this assumption are the generally high correlations among the spans of a

variety of S-R relations (Brener, 1940). These same data, paradoxically, provide a critical problem for the S-R_e model; the digit span test produces consistently higher estimates of ES than do other word span tests (Brener, 1940; Calhoon, 1935). The effect of materials on span is referred to as the materials effect.

Brener's study (1940) outlined the essential problem. He measured the spans for a variety of verbal materials (some of his span scores are presented in Table 1) and concluded that Ss can be conceived as having an underlying span ability which occurs in different amounts for different Ss. He concluded also that materials have an effect in addition to the basic subject variable; for example, digits and consonants produce conspicuously higher scores than other word tests.

TABLE 1.--Estimated elicitation spans produced by various classes of S-R units

Tests	Means
Digits	7.98
Consonants (visual presentation)	7.30
Consonants (oral presentation)	7.21
Concrete words (visual presentation)	5.76
Concrete words (oral presentation)	5.86
Abstract words (visual presentation)	5.24
Abstract words (oral presentation)	5.58
Colors (visual presentation of color samples)	7.06
Geometrical designs (visual presentation of designs)	5.31
Memory for commissions	2.42
	(2.42 x 3 =7.26)

Note.--This table was modified from Brener (1940). The Ss were 40 college students. The rate of presentation was one stimulus each 2 seconds.

There are two problems which must be dealt with in order to

resolve the materials effect in terms of the S-R_e model. First, a procedure for measuring the ES directly must be specified; second, the materials effect must be accounted for theoretically.

The "True" ES

Brener (1940) factor analyzed his data and reported the communality of each ES test (that part of the true variance of a given test due to factors common to other tests in the battery). The four largest communalities (1 is maximum) were .930, .908, .892, and .877, for abstract words, consonants, colors, and concrete nouns, respectively (in the case of colors, E presented strings of color samples, not printed or orally presented words; and S reported color names verbally). The communality figure for digits was .833 which ranked thirteenth out of 17 tests, suggesting variability in addition to that common to other tests. Brener also extracted a strong "memory span" factor. The loadings for abstract words, concrete words, and digits were .858, .869, and .858, respectively. If it is assumed that the factor common to all the tests is the ES ("memory span"), then the communality data indicate that abstract words and concrete nouns produce spans which are more closely related to ES than are digit spans. The loadings on the general factor indicate that all the spans are good measures of ES. The communality data and loadings on the general factor indicate that digit span reflects variation not related to ES.

Brener's results imply that the digit span is not the best measure of ES. Since the communality for abstract nouns was the highest of all communalities measured, abstract nouns would seem

to be the best estimate of ES. Brener did not define abstract nouns, but they were probably words such as health and beauty. These words would be relatively unfamiliar to retarded Ss because of the retardates' relative lack of education. For this reason, common words are likely to be better estimators of ES when retarded Ss are under consideration. Brener's estimate of span for concrete nouns was quite comparable to his estimate for abstract nouns (5.86 and 5.58, respectively). The communality of the concrete nouns test was .887 which ranked fourth in his total battery of tests. This evidence seems to be a reasonable justification for taking the results of word span tests (in which common one-syllable words are used) as direct estimators of ES.

Explanation of the Materials Effect

If the word span is the best estimate of ES, then the digit span is larger than the ES because the digit span is larger than the word span. Digit span may be larger than word span because the number of digits in a digit response string is larger than the number of responses in the same string. In other words, some digit responses may comprise two or more associated digits funtioning as one response. For example, the digit strings 8572 and 73295 (the underlining indicates an associated pair of digits) have the same number of responses but the first string has four digits; the second, five. This explanation of the materials effect is called the response size hypothesis (RSH) because responses which comprise two or more units are called large responses. The materials effect is attributed to a larger average response size in digit strings compared to word strings.

It is not unreasonable to expect greater numbers of large responses in digit strings compared with word strings because digits and words are used differently in normal conversation. The average person has an extensive history of repetition of almost every conceivable pair and triplet of digits (as, for example, repeating a telephone number: two-three-six-four-nine). This experience is the product of years of arithmetic and mathematics training, use of house numbers, test grades, ages, weights, heights, identification numbers, telephone numbers, and ZIP codes. In contrast to digits, a randomly selected group of common words is unlikely to contain pairs which frequently occur as pairs in normal usage. For example, hair, door, grass, and foot are all words used in this study. Any pair or triplet of these words is unlikely to occur as a pair or triplet in common conversation. Thus, on the average, S should have more large digit responses than large word responses in his S-R repertoire; and large responses imply a larger digit span.

Since verbal behavior may be an important variable controlling the development of large responses, age and IQ may be related to the materials effect. The CA directly measures the time S has had to engage in verbal behaviors of all types; therefore age, per se, is not related to differential practice effects for digits as compared with words. The IQ is closely related to school experiences and is a rough index of the quality and complexity of daily activity. Experience with digits should be grossly related to IQ because the use of numbers is related to level of education and complexity of daily activity. For these reasons, it was

hypothesized that the materials effect is directly related to IQ but unrelated to CA.

According to the RSH, the word span estimates the "true" ES; but the digit span reflects variation due to both ES and experience with digits. Based upon this interpretation, it was hypothesized that the variance of digit spans is greater than the variance of word spans.

Since the digit span and the word span each reflect the underlying ES, the digit span should never be smaller than the word span. It was hypothesized that digit spans are equal to or larger than word spans.

A note on terminology may prevent confusion later. The term response is being used with a special meaning. The response is a theoretical behavior unit (see Footnote 1) and may consist of one or more associated words or other verbal units such as phonemes. Response is always used in the theoretical sense. When reference is made to a word or digit as an observable behavior unit, then the terms used are word, digit, or the general term, unit. A unit may be larger or smaller than a response; a response may be larger or smaller than a unit. Units are defined by E and can be observed objectively, but responses must be inferred from these observations.

Direct Tests of the Response Size Hypothesis

The RSH attributes the materials effect to large responses in response strings. Direct evidence of the presence of large responses would strongly support the RSH, but large responses cannot be observed directly because they would be confused with groups of simple responses. Nevertheless, large responses should have

several specific effects on response strings which can be detected objectively and from which the presence of large responses may be inferred.

Transition numbers. The present discussion refers to the elicitation span profile (profile) shown in Figure 1. There are three distinct portions to the profile: the basal, the ceiling, and the transition portions. In the basal portion, S produces all strings perfectly on all trials. In the ceiling portion, S produces all strings incorrectly. In the transition portion, S produces strings of a given size both correctly and incorrectly. The number of string sizes in the transition portion is called the transition number (TN). The TN is an index of the variability or unpredictability of S's behavior near ES. The "ideal" S has a transition number of 0; his profile has no transition zone and his ES is the largest string in the basal portion. The typical S's behavior is not so predictable.

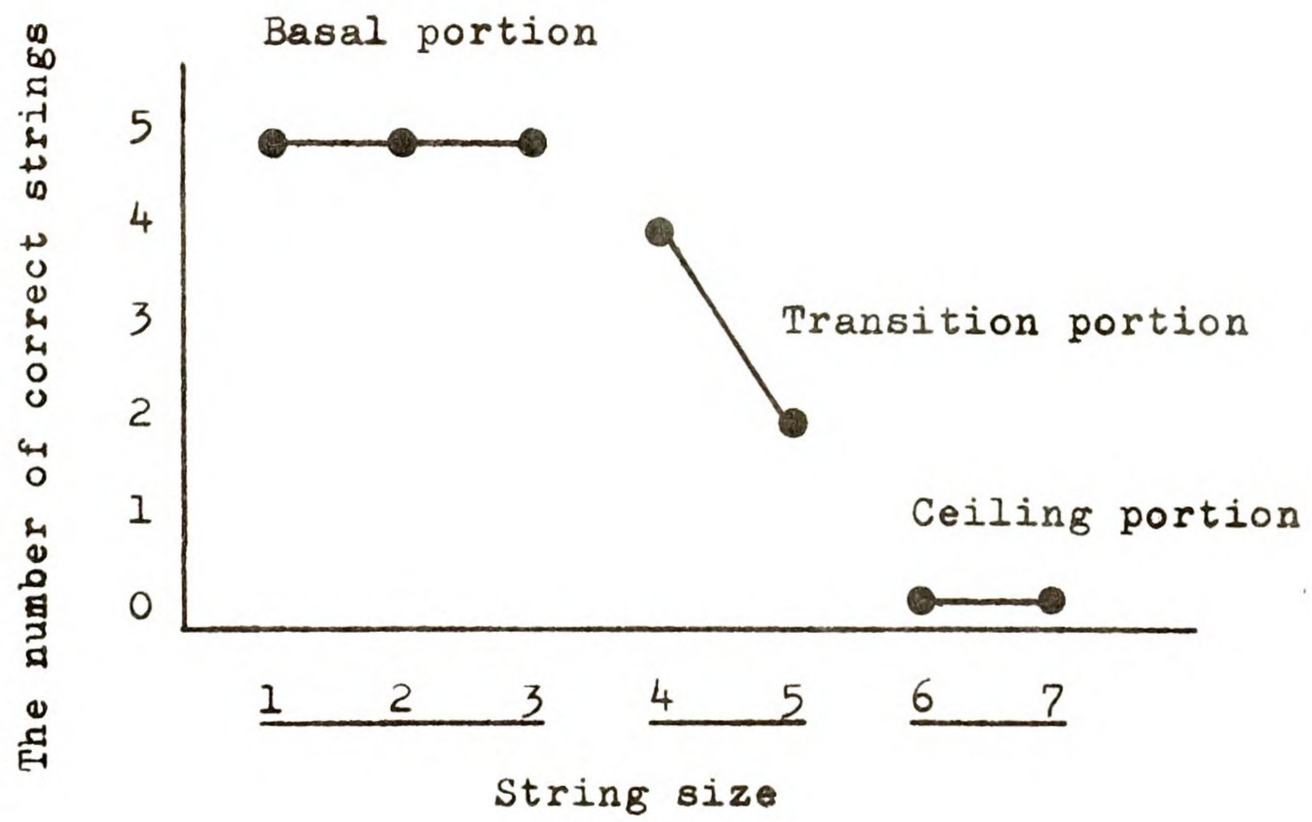

Fig. 1--A typical elicitation span profile. The transition number is two, the number of string sizes in the transition portion.

According to the RSH, two response strings may differ in the number of units and at the same time have the same number of responses (for example, 8572 and 73295). This variability of size of response string due to variability of response size should expand the transition zone of the profile because S would occasionally produce an atypically long string. It was hypothesized that digit TNs are larger than word TNs because digit strings have more large responses than do word strings.

The intrastring error rate. The intrastring (IS) error rate is the rate of errors in the second and penultimate positions of incorrect response strings which have a correct first or second verbal unit, respectively. The IS error rate is related to the presence of large responses because if the first or last verbal unit is correct and large, the second or penultimate verbal unit, respectively, must also be correct because it is associated to the first or last verbal unit. For example, in the string 536924 (the underline indicates an associated pair), the second unit, 3, has to be correct (if 5 is correct) because it is part of a response unit. The penultimate digit, 2, however, is independent of the final digit because each is an independent response.

Higher-order transpositions. A common error is the transposition of units (Conrad, 1959). Most transpositions are simple inversions of the order of two units; but a few transpositions are higher-order, involving three or four words. For example, the stimulus string 58734 might be reproduced as 58347. The transposition of 7 and 34 involves three digits so it is a higher-order transposition. Higher-order transpositions are assumed to

be transpositions involving large responses. It was hypothesized that digits produce more higher-order transpositions than do common words. The total rate of simple and higher-order transpositions is assumed to be the same for both words and digits.

Word-like Units

The S-R_e model attributes small digit and word spans to small ESs. It is possible, however, that small spans result from small responses just as large digit spans result, in part, from large responses. A small response would have to be smaller than a digit or word, that is, a phoneme. The transposition of phonemes should produce word-like units which do not occur in the stimulus string. As an example, consider the string car-boy-shoe. A transposition of phonemes could result in this response string: bar-coy-shoe. It was hypothesized that the number of word-like units is minimal (small spans do not result from small responses).

Table 2 summarizes all experimental hypotheses.

TABLE 2.--A summary of hypotheses

Classical phenomena
Digit span is positively correlated to IQ.
Word span is positively correlated to IQ.
Digit span is positively correlated to word span.
Digit span is larger than word span.
The response size hypothesis and the S-R_e model
Digit span is equal to or larger than word span.
The materials effect is positively correlated to IQ.
The variance of digit span is greater than that of word span.
Practice does not improve performance in digit or word span tests.
The rate of word-like units is minimal.
Direct tests of the response size hypothesis
Digits produce higher transition numbers than words.
Digits produce lower rates of intrastring errors than do words.
Digits produce more higher-order transpositions than do words.

Method

Subjects

Thirty[2] Ss were chosen from among the residents of Caro State Home and Training School, Caro, Michigan. Selection was aimed at a wide range of ES scores. Some Ss were individuals who were examined as part of the institutional testing schedule and for whom it was convenient to serve in this study. Other Ss were selected randomly from the high and low digit span groups of Experiment 2 so as to have Ss with a wide range of ES. Table 3

[2]Thirty-two Ss were actually tested, but two Ss were dropped because their digit scores were higher than 6. Since the largest string size of the test was seven, these two Ss made too few errors for proper analysis.

presents a statistical summary of the ages, IQs, digit spans, and word spans of the 30 Ss.

TABLE 3.--A statistical summary of the ages, IQs, digit spans, and word spans of the 30 Ss

Characteristic	Mean	Range	SD
CA (years)	25.1	17.5-40.5	6.39
IQ[a]	69.83	52.3-107.6	13.30
Digit span[b]	4.20	2-6	1.20
Word span[b]	3.57	2-5	.87

Note.--All Ss were epileptic and were taking anticonvulsants or tranquilizers at the time of testing.

[a]The IQ score is the average of all available tests the S had had in his lifetime. Only scores from the common tests were used: the Stanford-Binet, the Wechsler-Bellevue, the WISC, the WAIS, and the Peabody Picture Vocabulary Test.

[b]The digit and word spans are the absolute spans, that is, the size of the largest string which S produced correctly two or three times in three attempts.

Materials

Words and digits. The 10 digits and 10 familiar words were used to form strings for the two conditions, digits (D) and words (W). The familiar words were grass, soap, foot, tree, milk, light, door, phone, hair, and car. These words were chosen because they probably were familiar to most potential Ss.

The experimental lists. Each experimental list was made up of blocks of strings. Strings were generated by the random selection of words or digits with two restrictions: No word or digit was repeated in a string, and no word in the first or last

position of a string was repeated in the same position in the following string.

A block had either words or digits, never both. There were seven strings in a block, one of each size from one to seven units. The string sizes were randomly ordered within each block.

Each list contained three blocks of words and three blocks of digits. Two of the four lists presented the blocks in the order DWWDDW; the other two lists, WDDWWD. The four lists were distributed randomly among Ss.

In summary, each of the four lists presented both conditions. The lists differed in the specific strings presented, in the order of presentation of strings of different sizes, and in the order in which the two conditions were presented. The chief reason for making four lists was to counterbalance possible practice effects and to minimize the effects of specific strings which might be unusually easy, such as 5432. Such easy strings are usually eliminated, but in the present case the use of extra lists was preferred because string difficulty has been neither defined nor measured for the memory span. In addition, the easiness factor may be precisely what is being dealt with in the materials effect.

Procedure

When S arrived in the testing room E asked several questions and encouraged brief conversation. To ensure that each S could hear and produce each unit presented singly, E asked: "Would you please repeat each of these words after I say it." E then read the following words and allowed S to respond to each one: grass, soap, foot, tree, milk, light, door, phone, hair, car. E then

read these instructions: "I have a simple test here which will tell me how you remember words. Try to do the best you can but don't worry if you forget some because I put some very hard ones in which everyone will forget. I am going to read some words which I want you to repeat after me just the way I read them to you. Let's practice one: foot-car."

If S erred on this string E presented each of five two-word strings. If S produced three or more of these two-word strings correctly, he was accepted as an S; if not, he was dropped from the study. The five strings were grass-milk, light-hair, car-foot, tree-ball, and phone-grass. E then said, "That's the kind of thing we'll be doing. Try to do the best you can." At this point, E administered the ES test.

E read the words and digits at a rate of one per second. This rate was developed and maintained by practice with a clock. During presentation of a stimulus string, no timing device was used; but between trials E tested his rate. A 30 second interval was maintained between the end of one response string and the reading of the next string. An interval of this length was shown by Conrad (1960) to eliminate serial order intrusions from the preceding string without changing S's accuracy. It was necessary to prevent serial order intrusions because intrusions could often appear to be transpositions.

Results and Discussion

Classical Phenomena

The data reproduced the classical phenomena of the sequential auditory memory span: the positive relation between span

scores and intelligence, the positive relation between span scores measured by different S-R units, and the materials effect. The digit and word span scores by which these phenomena were measured were of two types: the absolute span (AB-DS, AB-WS) and the number of correct response strings (DS, WS). The absolute span (AB-DS or AB-WS) was defined as the largest string size which S produced correctly two or three times in three attempts. The number of correct strings (DS or WS), because of its simplicity and wide range of potential values (0 to 42), is a better dependent variable than the absolute span. The absolute span could assume only discrete values ranging from 0 to 7. The two measures of digit performance correlated .95 as did the two measures of word performance.

The number of correct strings is mathematically independent of the absolute span, but the measures are closely related statistically (r_s=.95) because S attempted an identical number of strings of each size. Ss with small ESs produced only the small strings correctly; Ss with large ESs produced both small and large strings correctly. It is important to note that the two measures provide separate indexes of ES for each S. AB-DS (an absolute measure) is not an estimate of DS (the number of correct strings); AB-DS is not a transformation of DS. Later, the two measures of performance are used as independent measures of ES for calculating partial correlations.

In the following presentation of results, the number of correct strings is used in all cases; and, occasionally, where relevant, the absolute measures are discussed. The two types of

measures correlated .95 so the absolute measure of performance rarely added information.

Digit span and IQ. It has been found repeatedly that digit span is related to IQ. Previous work has been done with groups consisting primarily of normal Ss (IQ standardization groups), but noncorrelational studies indicate that the digit spans of retarded Ss are smaller than those of normal Ss. The digit spans of the present Ss (primarily retarded and borderline intelligence) correlated .47 ($p<.01$) with IQ. The direction of the correlation is consistent with other data, but the size is lower than other reported correlations which are about .75. Inspection of the data revealed that two Ss were atypical and might adversely affect the correlation. The two Ss were nonretarded (IQs 104 and 107.6, the next lowest IQ was 86.5) and their digit spans were far lower than might be expected from a linear model of the relation between digit span and IQ (recall that a normal IQ with a low digit span is diagnostic of brain damage). The correlation between IQ and digit span excluding these atypical Ss was .56 ($p<.005$). This figure is still lower than .75 and may simply reflect a restricted range of IQs in the present sample (52.3-86.5, excluding the two atypical Ss), or boredom may have produced excessively variable span scores (sessions were 30 minutes long with a 30 second intertrial interval).

Word span and IQ. Previous studies have shown that the various measures of memory span are correlated, but usually only the digit span is compared with intelligence measures. The present data are perhaps the first to show that word span is also

positively correlated with IQ in retarded and low normal Ss, r=.36 (p<.05). As with digit span, the two "normal" Ss with IQs above 100 produced low word spans which were obviously discontinuous with the other data. The correlation between word span and IQ, excluding these two Ss, was .42 (p<.025).

Digit span and word span. It has been found repeatedly that the measures of span based upon various S-R classes are highly correlated; the present data were no exception. The correlation between DS and WS was .79 (p<.001) which was expected because both DS and WS measure ES. Ordinarily, one would expect an even higher correlation between two measures of the same quantity; but in this case the correlation is only .79 because, while digit span is assumed to measure both ES and experience with digits, word span measures only ES.

The materials effect. It has been found consistently that digit spans are larger than word spans among normal Ss (usually college students). The present data reproduced this common finding within a group of retarded and borderline normal Ss. The average absolute digit span was larger than the average absolute word span ($t=4.77$, $df=29$, $p<.001$), and the average number of correct digit strings was larger than the average number of correct word strings ($t=5.07$, $df=29$, $p<.001$). These data are discussed in more detail in the next section.

Phenomena Relevant to the Response Size Hypothesis

The materials effect. The materials effect was measured in two ways based on the two measures of digit and word span. The absolute materials effect (AB-ME) is the difference between S's

AB-DS and AB-WS (the absolute digit and word spans), and the ME is the difference between S's DS and WS (scores based on the number of correct strings). The AB-ME is a cruder measure than the ME because the AB-ME is unresponsive to small differences in digit and word performance. If, for example, S produced three correct four-digit strings but only two four-word strings, S would score an ME of 1. His AB-DS and AB-WS, however, would be equal; so his AB-ME would be 0. The two measures of the materials effect correlated .81 ($p<.001$).

It was hypothesized that digit spans are equal to or larger than word spans; that is, the materials effect ranges upwards from 0. The AB-ME ranged from 0 to 2 ($\bar{X}=.63$, $SD=.71$). The ME ranged from -2 to 6 ($\overline{ME}=1.9$, $SD=2.02$). The $\overline{ME}$ was significantly larger than 0 ($t=5.14$, $df=29$, $p<.001$). WS was larger than DS for only 2 Ss out of 30 and the difference in each case was small, -1 and -2 (the AB-ME scores of these 2 Ss were 0). An ME of -2 is a small difference between digit and word performance because a difference of one between the absolute digit span and the absolute word span would produce an average ME of -3. Only 2 Ss out of 30 accurately produced a word string larger than his largest correct digit string. The evidence is clear that the materials effect works only in the direction of greater digit scores.

As hypothesized, the materials effect (ME) was positively correlated with IQ, $r=.36$ ($p<.01$). The two Ss with normal IQs had atypical materials effects; the correlation excluding these Ss was .44 ($p<.01$). As hypothesized, the correlation between age and the materials effect (ME) was small and nonsignificant, $r=.23$ ($p>.10$).

As hypothesized, digit span and the materials effect (DS and ME) were correlated, r=.74 (p<.001).

The word span is taken as a "true" measure of ES which does not reflect the digit experience which produces the materials effect; word span and the materials effect should not be related. Word span (WS) and ME were nonsignificantly correlated, r=.16 (p=.20).

Since the materials effect was measured as the difference between the two conditions, digits and words, the relevant data were summarized by analysis of variance which revealed the materials effect as a significant materials factor. The effects of age, IQ, and ES on the materials effect, as revealed by analyses of varience, are discussed briefly in the context of the practice effect below.

The variances of digit and word spans. As hypothesized, digit span was more variable than word span. The variance of AB-DS was greater than the variance of AB-WS (F=1.91, df=29/29, p<.05); the variance of DS was greater than the variance of WS (F=2.10, df=29/29, p<.025).

Word-like units. Word-like units did not occur. This is taken as evidence that the responses produced by all Ss (with the exception of the large responses which produced the materials effect) are of comparable size and that individual differences in digit and word spans reflect true differences in ES.

A Summary and Interpretation

The S-R_e model and the RSH suggest the following statements of the relations between ES, digit span, word span, materials

effect, and IQ. These formulas are not post hoc models; each embodies formally the concepts and assumptions implicit in the design of this experiment.

(1) Word span = f(ES)	The word span is taken as a true measure of ES.
(2) Digit span = f(ES) + f(EXd)	Digit span measures ES, but digit span also reflects differences in experience with digits (EXd). EXd is related to, but not equivalent to, EX (IQ-related experiences such as schooling, use of digits, and social experiences and responsibilities).
(3) IQ = f(ES) + f(EX)	The IQ reflects ES because a test of ES is frequently part of the IQ test and because differences in ES contribute to the effectiveness of schooling. IQ is primarily a measure of achievement and reflects EX (IQ-related experiences such as schooling, use of digits, and social experiences and responsibilities).
(4) The materials effect = f(EXd)	According to the RSH, the materials effect is due to specific experience with digits. This experience is related indirectly to EX.

Formulas 1-4 predict the various simple and partial correlations produced by the present data. Word span should be correlated with digit span and IQ because all three variables are measures of ES (formulas 1, 2, and 3). All these correlations were significant as discussed earlier. Word span and the materials effect should not be correlated because each measures different and unrelated quantities (formulas 1 and 4). These correlations were small and nonsignificant (.13 to .24).

The technique of partial correlation allows examination of the correlation between two variables while statistically holding a third variable constant. For example, if digit span and word span are correlated simply because each is correlated with a third variable, say IQ, then, by holding IQ constant, the correlation between digit span and word span should be reduced or eliminated. The correlations between digit span and word span ranged from .78 to .80. When IQ was held constant, the following correlations resulted:

$$r_{AB\text{-}DS,\ AB\text{-}WS\ .\ IQ} = .78\ (p<.001)$$

$$r_{DS,\ WS\ .\ IQ} = .75\ (p<.001)$$

$$r_{DS,\ AB\text{-}WS\ .\ IQ} = .78\ (p<.001)$$

$$r_{AB\text{-}DS,\ WS\ .\ IQ} = .83\ (p<.001)$$

These partial correlations indicate that the correlation between digit span and word span stems not from their common correlation to the IQ; but from the fact that digit span, word span, and IQ all reflect variations in ES (formulas 1, 2, and 3).

The relatively low correlations (.31 and .36) between word span and IQ imply that the IQ is not a good measure of ES and is chiefly a measure of EX (IQ-related experience). The small relation between word span and IQ should be reduced or eliminated if the variations in ES are held constant. To hold ES constant, a measure of ES must be held constant in the partial correlation procedure. The word span, digit span, and the IQ all reflect ES (formulas 1, 2, and 3). Obviously, IQ cannot be used because it would enter twice into the computations. AB-DS, DS, AB-WS, and WS provide independent measures of ES; each can serve as the control

variable so long as it does not occur twice in the calculations. The following correlations estimate the relation between word span and IQ with ES partialed out:

$$r_{WS,\ IQ\ .\ DS} = -.02\ (p=.54)$$

$$r_{WS,\ IQ\ .\ AB\text{-}WS} = .22\ (p=.13)$$

$$r_{WS,\ IQ\ .\ AB\text{-}DS} = .13\ (p=.25)$$

In contrast to word span, the digit span should be related to IQ even if ES is held constant because the digit span and IQ each measure variation due to IQ-related experience (EXd and EX, formulas 2 and 3). The following correlations measure the relation between digit span and IQ with ES held constant:

$$r_{DS,\ IQ\ .\ WS} = .32\ (p=.05)$$

$$r_{DS,\ IQ\ .\ AB\text{-}WS} = .39\ (p=.02)$$

$$r_{DS,\ IQ\ .\ AB\text{-}DS} = .41\ (p=.01)$$

The previous correlations show that the digit span is related to IQ even though the ES is held constant. The relation should hold as well even when EX is held constant (formulas 1 and 2). According to formula 4, the materials effect measures EXd which is related to EX since both EXd and EX are IQ-related variables. The following correlations show the relation between digit span and IQ with the variation due to IQ-related experiences held constant:

$$r_{DS,\ IQ\ .\ ME} = .33\ (p<.04)$$

$$r_{DS,\ IQ\ .\ AB\text{-}ME} = .42\ (p=.01)$$

The materials effect is correlated with IQ because both measures reflect variance due to IQ-related experiences (formulas 3 and 4). If EX and EXd are controlled as sources of variance, the

correlation between the materials effect and IQ should be reduced or eliminated. The experience factor can be controlled by holding digit span constant because the formulas indicate that the correlations between digit span and materials effect are due to variation in experience:

$$r_{ME,\ IQ\ .\ DS} = .02\ (p = .46)$$

$$r_{ME,\ IQ\ .\ AB\text{-}DS} = .16\ (p = .21)$$

Similarly, if the variation in ES is partialed out, the materials effect should still be correlated with IQ (formulas 3 and 4). The word span is the best measure of ES and serves as the control variable in the following correlations:

$$r_{ME,\ IQ\ .\ WS} = .31\ (p = .05)$$

$$r_{ME,\ IQ\ .\ AB\text{-}WS} = .31\ (p = .05)$$

The correlations between digit span and the materials effect are high, .62 to .74. They should be even higher if the variance due to ES is controlled (formulas 2 and 4). The AB-WS was taken as the best estimate of ES and used as the control variable:

$$r_{ME,\ DS\ .\ AB\text{-}WS} = .94\ (p < .001)$$

$$r_{ME,\ AB\text{-}DS\ .\ AB\text{-}WS} = .80\ (p < .001)$$

In summary, even though the practice of computing large numbers of correlations from the same data carries a high risk of type I error, the consistency of the results with the predictions lends considerable credence to both the correlations and the formulas.

The Effect of Practice

The experimental lists were constructed so as to measure the effects of practice on both digit and word strings. Each S

attempted six blocks of strings, three blocks of word strings and three blocks of digit strings. Each block contained one string of each size, one through seven units. These blocks were presented in a counterbalanced order so that for each S the total session consisted of three stages of practice and each stage of practice consisted of one block of digit strings and one block of word strings. Over all Ss, each stage consisted of equal numbers of digit and word strings of each size counterbalanced for order of presentation within each stage. Four different lists were assigned at random to Ss. The lists variable was not expected to have an important effect; it was not crossed with other variables and was not extracted in the analysis.

The same set of data (number of correct strings) was organized in four different ways to assess the effects of each of four subject variables: IQ, age, elicitation span, and the materials effect. On the basis of each subject variable, the 30 Ss were divided into 15 high Ss and 15 low Ss. The data were analyzed by means of 2 X 2 X 3 analyses of variance (high and low subject grouping X digits and words X three stages of practice).

As expected, there was no improvement with practice; but there was a significant decline of performance with practice ($p<.01$ in all four analyses, see Figure 2). This decline in performance was unexpected and may have resulted from a general boredom effect, a response habituation effect (repeated elicitation of a response results in a reduction of response probability), or a proactive inhibition effect (forgetting is a function of the amount previously learned). This latter explanation is unlikely because

proactive inhibition in short-term memory tasks has been found only at long delay intervals (18 to 27 seconds, Keppel & Underwood, 1962; Loess, 1964). The response habituation hypothesis suggests that the effect of practice is stimulus specific; that is, the decline in performance should be proportional to the number of presentations of each stimulus. Also, the effect of practice should not transfer to stimuli which have been presented infrequently even though all stimuli may belong to a common class. The boredom interpretation implies that the decline is general; that is, the decline in performance affects all stimuli regardless of the number of presentations.

The materials factor was significant in all analyses ($p<.01$ in each case, see Figure 2) which simply reflects the common finding that digit spans are larger than word spans. Previous studies dealt with normal Ss; the present data extend the principle to retarded and borderline normal Ss. Practice did not interact with materials (Fs=.90 to .92 in the four analyses). This finding means that practice produced similar effects for both words and digits and rules out an explanation of the effect of practice in terms of habituation of large responses (large responses occur primarily in digit strings).

Practice, materials, and IQ. Practice did not interact with IQ; that is, both IQ groups responded similarly to practice (F=.35, df=2/56). The overall results of the analysis were not qualified by evidence of a Materials X IQ X Practice interaction (F=1.18, df=2/56). The findings are summarized in Figure 2; the analysis of variance, in Table 4.

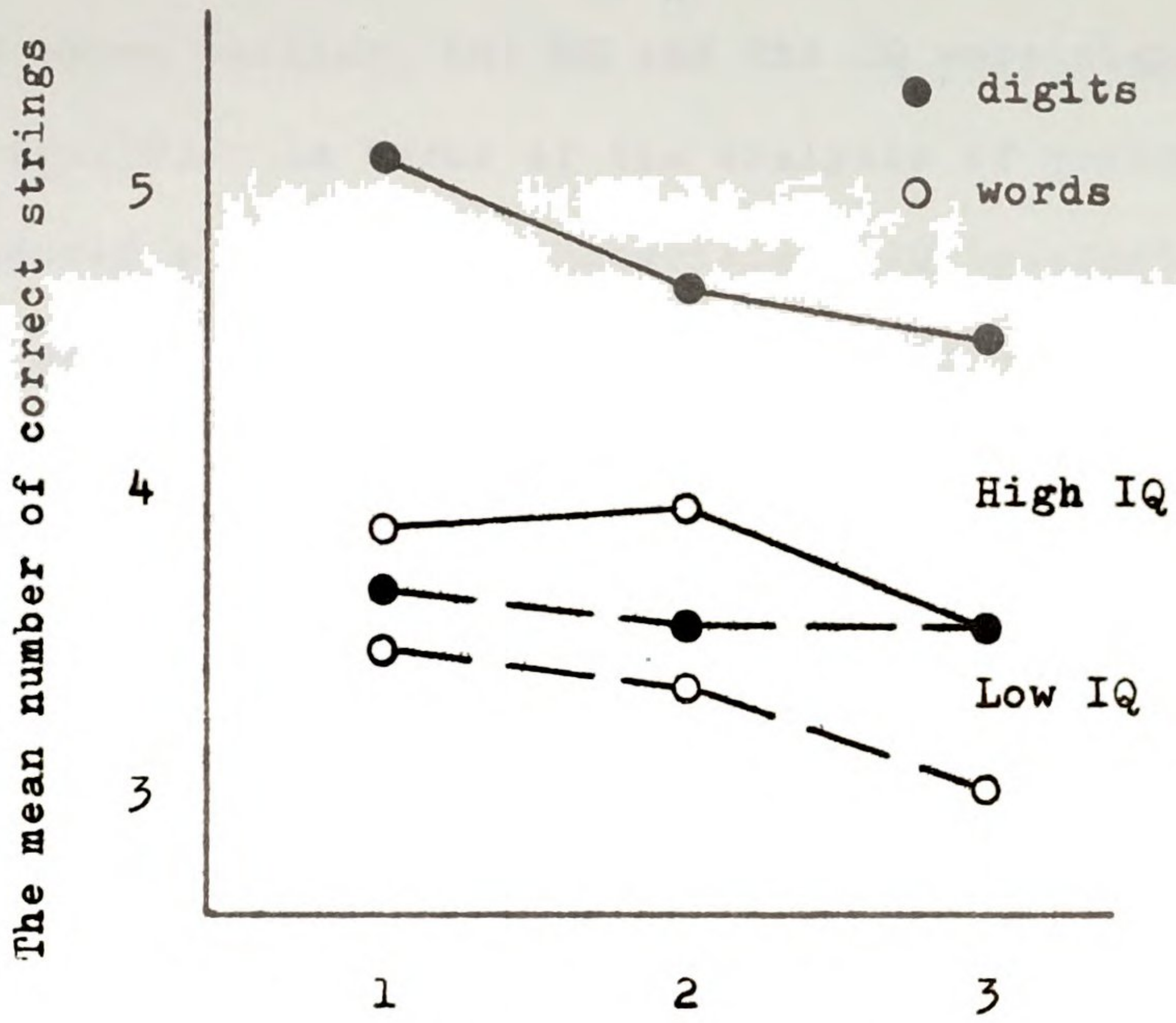

Fig. 2.--The effects of practice, IQ level, and materials on elicitation span performance.

TABLE 4.--A summary of the analysis of variance on the number of correct response strings as a function of IQ, materials, and practice

Source	df	MS	F
Between Subjects			
IQ	1	32.08	8.65 ***
Ss within groups	28	3.71	
Within Subjects			
Materials	1	18.68	32.77 ***
Materials X IQ	1	5.01	8.79 ***
Materials X Ss within groups	28	.57	
Practice	2	2.42	5.26 ***
IQ X Practice	2	.16	.35
Practice X Ss within groups	58	.46	
Materials X Practice	2	.36	.92
Materials X IQ X Practice	2	.46	1.18
Materials X Practice X Ss within groups	56	.39	

*** $p < .01$

As shown earlier, the ME and the IQ were significantly correlated (r=.36). In terms of the analysis of variance, the same data produced a significant Materials X IQ interaction (F=8.79, df=1/56, p<.01). High IQ Ss had larger differences between digit and word scores than did low IQ Ss.

Practice, materials, and ES. According to the present interpretation, the ES is best estimated by word span rather than digit span. Nevertheless, this interpretation is not generally established; so it was felt best to divide Ss according to a general "memory span" score. The error would be small because digit span and word span are highly correlated (r=.79). S's level of ES was estimated by his combined digit and word performance, that is, the total number of correct strings.

Both ES groups responded similarly to practice; that is, the ES X Practice interaction did not approach significance (F=.67, df=2/56). The overall results of the analysis were not qualified by evidence of a Materials X ES X Practice interaction (F=.72, df=2/56).

Since Ss were divided on the basis of the number of correct strings, the main effect of ES was significant but trivial. As was shown earlier, digit span correlated about .70 with the materials effect; that is, large digit spans were associated with large differences between digit and word spans. Since the present subject division was based in part on the digit span, the interaction of Materials X ES was significant (F=10.36, df=1/56, p<.01). High ES subjects had larger differences between digit and word scores than did low ES subjects.

Practice, materials, and the materials effect. The size of S's materials effect was measured by the ME (the number of correct digit strings minus the number of correct word strings). The Practice X ME interaction was of borderline significance (F=3.00, df=2/56, p<.10). The lower ME group showed only a small effect of practice.

This result was not predicted and there is no obvious theoretical interpretation; the effect is probably a floor effect. The scores of the Ss in the low ME group approached a mean of three correct strings for each material. Whereas S could score as low as 0, he could score 3 simply by producing each one-unit string in each block of strings. One-unit strings were very easy for the present Ss (mean word span =3.57) so the floor effect interpretation is not unreasonable.

Another interpretation was not supported by other data. Since dividing Ss according to the materials effect presumably divides Ss according to the presence of large responses in their response strings, it would appear that the practice effect was restricted to Ss who produced large responses. This interpretation has a certain plausibility because response habituation effects would be expected to be more pronounced for the larger responses which presumably are weaker than the primary simple responses in the present situation. The problem with this interpretation is that the word strings are presumed to have few large responses; yet in this and the other analyses, practice had a similar effect on both word and digit performance. The results of this analysis were not qualified by evidence of a Materials X Practice

X ME interaction (F=.72, df=2/36).

It was shown earlier that ME correlated .74 with digit span; that is, large materials effects were associated with large digit scores. The result in the present analysis was a significant main effect of ME (F=13.85, df=1/56, p<.01). The Materials X ME interaction was significant but trivial because the division of Ss by ME was based on the differences between digit and word spans.

Practice, materials, and age. The Age X Practice interaction did not approach significance (F=.64, df=2/56). This result indicates that both age groups responded similarly to practice. The overall results of the analysis were not qualified by evidence of a Materials X Age X Practice interaction (F=.20, df=2/56).

As discussed earlier, the literature is uniform in suggesting that the ES reaches an asymptote during adolescence; although there is some evidence that slight increases occur to age 25. Since the present Ss were adults, age was not expected to be related to performance. The higher age group had higher span scores, but the effect was not significant (F=1.75, df=1/56, p>.10). The slight evidence of an age effect probably reflects the slight growth of ES to age 25. The weak effect of age cannot be explained in terms of the IQ because the mean IQs of the older and younger groups were 68.46 and 71.20, respectively.

It was shown earlier that age does not correlate with the materials effect; that is, age is unrelated to the difference between digit and word performance (r=.23, p>.10). This fact is reflected in a nonsignificant Materials X Age interaction (F=.49, df=1/56).

Direct Tests of the Response Size Hypothesis

The present analysis of ME and AB-ME scores indicated that Ss differ in the extent to which they show the materials effect; 9 Ss did not show the materials effect to even a minimal extent (ME=0). For this reason, Ss were divided into two groups on the basis of size of ME scores. The division was made by ranking the 30 Ss according to the ME score then dividing the Ss into a top 15 and a lower 15 Ss. This division was successful in isolating the materials effect in the higher group; the mean ME scores of the two groups were 3.60 and .20. An ME of 1 means that the number of correct digit strings produced was one more than the number of correct word strings (equal numbers of digit and word strings were presented). An ME of .20 is a small effect because a difference of just one ES unit between the absolute digit span and the absolute word span would result in an ME of 3 or 4. Table 5 presents a summary of the subject characteristics of the two groups of Ss.

Transition numbers. The S-R_e model implies that S should perform perfectly on all strings up through ES and incorrectly on all strings beyond ES. This type of performance produces a TN of 0. Seven Ss in this study earned TNs of 0. It was hypothesized that TNs for digits are larger than TNs for words. Since the Ss were divided into two groups, with and without the materials effect, the hypothesis predicts a Materials Effect X Materials interaction such that Ss in the higher ME group have larger digit TNs than word TNs; but Ss in the lower ME group should show no difference in TNs as a function of materials. The data (Figure 3) were consistent with the prediction although the analysis of variance did not produce significant Fs (Table 6). The critical comparison, however,

was between the digit and word conditions within the high ME group. This difference was significant (matched pairs t=1.82, df=14, p<.05). Within the low ME group, materials had no effect (matched pairs t=.24, df=14, p>.40).

TABLE 5.--A summary of the ages, IQs, digit spans, and word spans of Ss with high materials effects and low materials effects (ME)

Characteristic	High ME	Low ME
Age (years)	$\overline{CA}$=27.30 (17.5-40.5) SD=6.94	$\overline{CA}$=22.90 (18.0-38.5) SD=4.88
IQ	$\overline{IQ}$=73.39 (52.3-107.6) SD=14.0	$\overline{IQ}$=66.27 (53.5-104.0) SD=11.31
Absolute digit span	$\overline{AB\text{-}DS}$=4.93 (3-6) SD=.87	$\overline{AB\text{-}DS}$=3.47 (2-5) SD=1.01
Absolute word span	$\overline{AB\text{-}WS}$=3.87 (3-5) SD=.79	$\overline{AB\text{-}WS}$=3.27 (2-5) SD=.89
Materials effect (ME)	$\overline{ME}$=3.60 (1-6) SD=1.31	$\overline{ME}$=.20 (0-1) SD=.83
Absolute materials effect (AB-ME)	$\overline{AB\text{-}ME}$=1.07 (0-2) SD=.67	$\overline{AB\text{-}ME}$=.20 (0-1) SD=.40

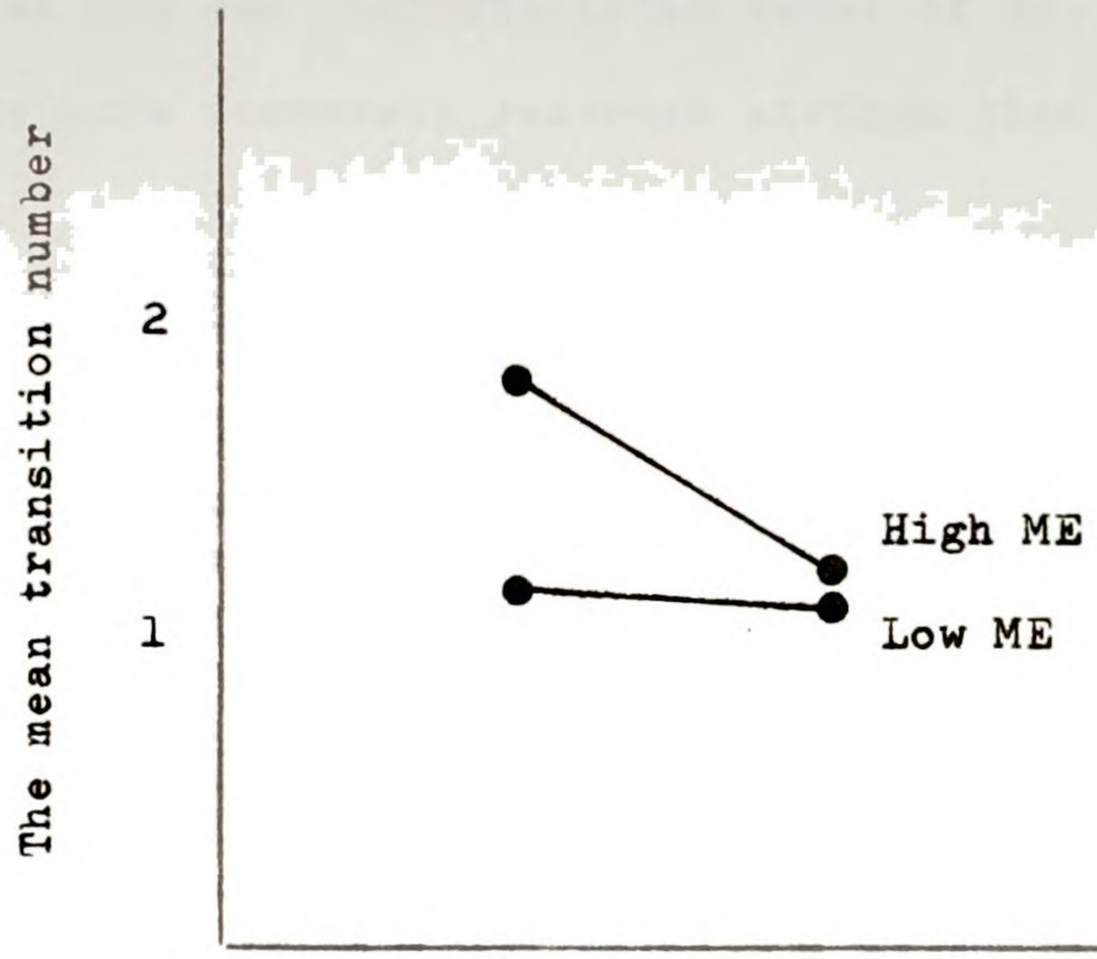

Fig. 3.--The effects of materials and the materials effect on transition numbers.

TABLE 6.--A summary of the analysis of variance on the transition numbers as a function of materials and the materials effect (ME)

Source	df	MS	F
Between subjects			
Level of ME	1	2.40	2.89 *
Ss within groups	28	.83	
Within subjects			
Materials	1	1.67	1.78 ns
ME X Materials	1	1.06	1.13
Materials X Ss within groups	28	.94	

ns $p > .10$

* $p < .10$

The IS errors. The datum used for the analysis was the rate of IS errors: the number of IS errors divided by the number of incorrect response strings. This datum was used because each S

performed at his own characteristic level of ES; subjects with low-ESs produce more incorrect response strings than subjects with higher ESs.

It was hypothesized that digits produce lower IS error rates than do words. Just as for the TNs, the effect should occur only with the materials effect, that is, for Ss with high ME scores under the digit condition. As predicted, the Materials X Materials Effect interaction (Figure 4) was significant (F=7.50, df=1/28, p<.05, Table 7). Within the low ME group, the difference between the digit and word conditions was not significant (matched pairs t=.89, df=14, p=.20).

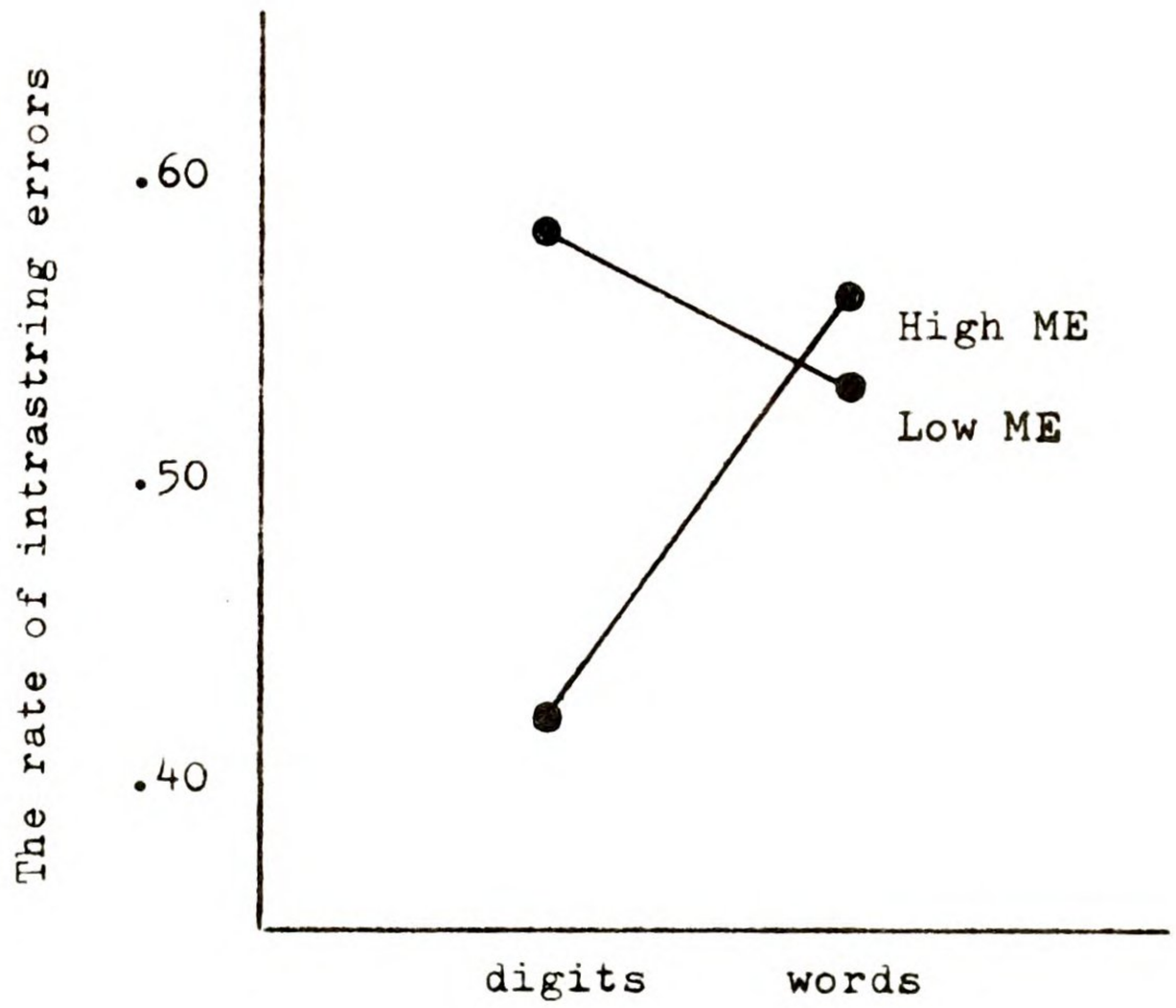

Fig. 4.--The effects of materials and the materials effect on the rate of intrastring errors.

TABLE 7.--A summary of the analysis of variance of the intrastring error rates as a function of materials and the materials effect (ME)

Source	df	MS	F
Between subjects			
Level of ME	1	.05	1.7 ns
Subjects within groups	28	.03	
Within subjects			
Material	1	.03	1.50 ns
Material X ME	1	.15	7.50 **
Material X Ss within groups	28	.02	

ns $p > .10$

** $p < .05$

Transpositions. In general, the analyses of the transposition data were disappointing because the data were so variable that the Fs were small in all cases (rates ranged from 0 to 1.0). The largest F had a probability greater than .25. The variability was due in part to limited opportunity to make transposition errors which occurred at low rates. The difficulty was compounded by the fact that Ss with high span scores, such as 5, had limited opportunity to make errors of any kind because the list was of constant composition for all Ss. Transpositions could be better studied by presenting Ss with string sizes at, and just beyond, span so as to produce large numbers of errors.

The significance of transpositions is unclear either with respect to the S-R_e model or the several memory span models (for example, Brown, 1958; Ellis; 1970). A transposition may be a transposition of responses or may be the result of intrusions which simply look like "transpositions." It was assumed that

transpositions could be taken at face value and be considered to reflect a process which somehow transposed responses in S's behavior. From this point of view, the higher-order transposition is a transposition involving a large response; and it was hypothesized that digits produce more higher-order transpositions than do word strings. The datum used for analysis was the rate of higher-order transpositions: the number of higher-order transpositions divided by the number of erroneous response strings. The rate datum was used rather than the number of higher-order transpositions because each S performed at a characteristic level of accuracy.

In terms of the analysis of variance, the present hypothesis predicts a significant interaction of Materials X Materials Effect: Within the high ME group, digits produce higher rates of higher-order transpositions than do words; within the lower ME group, materials have no effect.

The results (of poor statistical reliability) were exactly opposite the predictions (Figure 5). Within the high ME group there was no effect of materials; but within the low ME group, words produced higher rates of higher-order transpositions. The difference within the lower ME group was of borderline significance (matched pairs $t=1.31$, $df=14$, $p=.10$). The meaning of these data is not clear. The results are probably best attributed to chance variation pending a more adequate research design.

The data on higher-order transpositions certainly do not support the RSH hypothesis, but are not serious negative evidence for two reasons: The variability of the data are simply too great

to draw even tentative conclusions, and the meaning and significance of transposition errors have yet to be determined. Future research should be directed toward determining whether transpositions are actually transpositions of responses or simply intrusions which look like transpositions. Future attempts to measure the rates of transpositions must ensure the production of larger numbers of such errors so that stable rates can be determined.

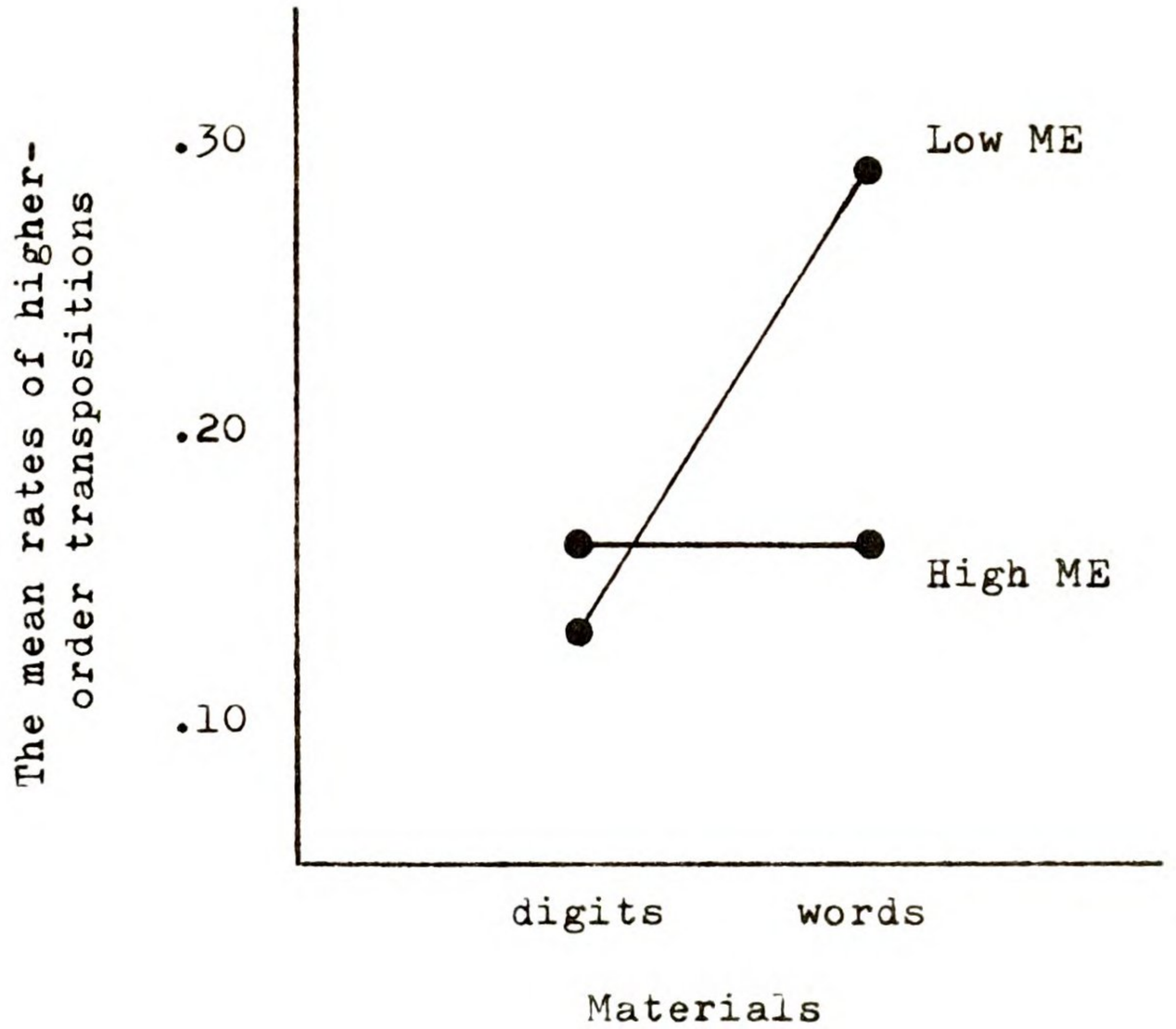

Fig. 5.--The effects of materials and the materials effect on the rate of higher-order transpositions.

It was assumed that the total transposition rate (simple plus higher-order) is equivalent for digits and words. In fact, digits produced higher mean rates of transpositions (.52 vs. .44); but the statistical tests did not approach significance. The largest difference between digits and words was in the high ME group, yet the difference did not approach significance (matched pairs t=.39,

df=14, p>.25). No valid conclusions can be drawn from these data.

Simple transpositions produced no better data than the higher-order transpositions. Digits produced higher rates of simple transpositions than did words (.45 vs. .37), and the effect of materials was greater for the low ME Ss. No overall effects approached significance. The digit and word conditions within each ME group were compared by matched pair ts neither of which approached significance: Within the high ME group, t=.58, df=14, p>.25; and within the low ME group, t=1.05, df=14, p>.10. There is no obvious theoretical reason for the results so it is best to suspend judgment pending an adequate research design.

In summary, practice produced a steady decline of performance within the experimental session. This conclusion was not qualified by interaction of practice with age, IQ, ES, or the size of the materials effect. There was nonsignificant evidence of an interaction of practice with the materials effect, but the interaction was attributed to a floor effect. The decline of performance with practice was attributed to boredom or to response habituation.

Experiment 2:

The Effects of Rate, Grouping, and Practice on the Digit Spans of Borderline and Retarded Adults

In the comparative approach to the analysis of the behavior of retardates, the best design seems to be the treatment X levels design in which Ss of different levels of IQ are compared in their response to experimental manipulation of stimulus variables (Baumeister, 1967). The desired finding is an interaction between

levels and treatments, that is, evidence that normal and retarded Ss respond differently to similar experimental manipulations.

Once differences between normal and retarded Ss are established, experimentation can take a different approach. Rather than continue looking for more differences, we can begin to try to show the sufficiency of a theoretical conception of a well demonstrated difference. This was the goal of the present study. It was assumed that a deficit in memory span is characteristic of at least some retardates. The S-R_e model formalizes the memory span deficit in terms of size of ES. How adequately can the S-R_e model account for the results of a treatment X levels design in which Ss are grouped according to their ES? This question was examined in terms of the effect of stimulus grouping and the effect of rate of presentation of stimuli on digit span performance in Ss of a wide range of ES.

The sufficiency principle. According to the S-R_e model, there are just two general types of individual differences: the size of ES and the composition of the S-R pool. In order to test the sufficiency of the S-R_e model to account for digit performance, both ES and the S-R pool must be controlled and their effects separated from those of the independent variables. Size of ES can be controlled by arranging Ss according to size of digit span which reflects differences in ES. The individual characteristics of the S-R pool can be controlled by using S-R units which all Ss have in comparable strength. Highly familiar words and digits were assumed adequate for this purpose. Within this scheme, the effects of experimental variables must be conceived in terms of the S-R

units elicited in the experiment because ES is assumed constant. The sufficiency principle implies that, provided a variable does not introduce unwanted responses in one group of Ss compared to another, then all Ss respond similarly to experimental manipulation, regardless of size of ES. Thus, in the treatments X ES design, interaction of treatments and ES levels are an indication that either the S-R_e model is in error (subjects differ in other ways than size of ES and S-R pool characteristics) or the target stimulus-response units were poorly controlled. A theoretical analysis of the results of such an experiment requires that special attention be paid to the possibility of contamination of results due to elicitation of unanticipated responses.

Rate of Stimulus Presentation

At extremely rapid rates of stimulus presentation, S would not perceive discrete stimuli. At extremely slow rates, intervening stimulation would become relatively more important than the stimulus string in controlling S's behavior. For example, a thorough description of S's behavior during an experimental session in which digit presentation occurred at a rate of one digit every 5 minutes would require a description of S's activities during the interstimulus interval, that is, S's responses to his surroundings. The present study was concerned with a range of rate variations slow enough to allow discrimination of stimuli and fast enough to be the predominant controlling stimuli (one-half, one, two, and four digits per second).

In the simplest hypothetical situation, each stimulus elicits just one response and this response is elicited immediately and

consistently for all Ss regardless of size of ES; that is, the rate of stimulus presentation does not interact with level of ES. So long as all Ss respond to each stimulus at rapid rates and so long as external stimuli do not compete for control at slow rates, each S performs at his characteristic level of ES. With respect to the treatments X ES design, the rate factor would produce no main effect and there would be no interaction of rate with ES levels.

Stimulus Spacing

The most common method of measuring digit span uses a regular rate of stimulus presentation, usually one digit per second (for example, Wechsler, 1955). There is evidence that grouping of digits improves performance (Spitz, 1966; MacMillan, 1970).

It was hypothesized that grouping has a positive effect on performance because of a practice effect. My pilot work indicated that many Ss "echo" the stimuli overtly during the presentation of the stimulus string; it is not unreasonable to assume that all or most Ss echo implicitly. In the ungrouped condition Ss would echo single digits; but in a grouped condition, pairs.

Recall from the general introduction that repetition of specific response sequences improves performance on the same sequence in subsequent strings. In the grouped condition, echoing results in practice of digit pairs so that subsequent production of the complete response string should be facilitated. In other words, "echoing" should develop large responses which then facilitate production of the response string as outlined in Experiment 1. Echoing of single digits should not produce an improvement because

the practice involved amounts to simple repetition of random digits which Melton (1963) and Hebb (1961) found to have no beneficial effect on performance.

There is no reason to suppose that the effects of "echoing" or that "echoing" itself are restricted to any one level of ES; therefore, it was hypothesized that grouping facilitates the performance of all Ss regardless of ES level. In other words, grouping was not expected to interact with levels of ES.

In summary, if E could adequately control the S-R units which occur in his study and if Ss could be divided accurately into ES levels, the results of the rate and grouping factors would be simple: Rate would have no important effect for Ss of all ES levels, and grouping would uniformly facilitate the performance of all Ss with ESs of at least 2 (an ES of 2 is necessary to echo pairs). Separating Ss according to ES is no problem theoretically; digit and word spans are closely related to ES. The problem of controlling the exact S-R events which occur is more difficult.

The Problem of Uncontrolled Large Responses

According to the response size hypothesis tested in Experiment 1, the materials effect results from the presence of large responses in S's response strings. The materials effect did not occur uniformly across Ss and was, in fact, related to IQ. This situation is interpreted to mean that, even though the stimuli presented to all Ss were identical; the responses elicited were not. Some Ss responded with large responses to a situation which produced only simple responses in others. This is not a surprising situation; probably all stimuli are associated with a variety of responses.

White (1965) theorized that a stimulus elicits several responses in succession. In other words, responses are "temporally stacked" in such a way that the exact response elicited by the stimulus varies as a function of the time between the stimulus presentation and the occurrence of the response. White presented specific data which show that the concept of temporal stacking is valid over a time interval of .3 seconds to 1.4 seconds, an interval which is comparable to those used in this study (one-half, one, two, and four digits each second).

If a stimulus is considered to elicit a time-based succession of responses, then the effects of rate and grouping must be reconsidered. Slow rates would obviously allow more responses to be elicited by a given stimulus. These responses could facilitate performance if they formed part of the target string. It is not unlikely that successive digit responses elicited by a stimulus would have a high probability of being part of the target string simply because there are only 10 digits. As an example, consider the stimulus string 47196. Each of the stimuli elicits the single spoken digit; but, according to the "temporal stacking" conception, each stimulus also is likely to elicit other responses such as other single digits or other pairs and triplets of digits (large responses). If the successive responses formed part of the target string, S's performance would be facilitated. If, in the illustrative string above, 7 elicited the verbal response one-nine, then S's performance would be facilitated. Note that these facilitation effects are related conceptually to the materials effect because both effects are based on the presence of large responses in S's

S-R pool. Note also that "thirty-nine" is not an example of the large response, three-nine.

Slow rates of stimulus presentation should allow each stimulus to elicit large responses; if S has the appropriate large responses in his S-R pool, his performance should benefit. High IQ Ss should respond positively to slow stimulus rates because high IQ Ss produce large digit responses. Low IQ Ss, in contrast, should not show this effect of rate because their S-R repertoires do not include the large responses; that is, they did not show a strong materials effect in Experiment 1.

Table 8 presents a summary of experimental hypotheses.

TABLE 8.--A summary of experimental hypotheses

Item	Hypotheses
Classical phenomena . . .	Digit span and IQ are positively correlated.
	Digit span and age are not correlated.
Practice	Practice produces a general decline in performance within a session.
	Repeated testing has no effect on digit span.
Rate	Slow rates increase digit spans of high IQ Ss.
	Rate has no effect on the digit spans of low IQ Ss.
Grouping	Grouping increases digit spans.
Rate X Grouping	No interaction is predicted.
ES levels	Neither rate nor grouping interacts with ES level.

Method

Subjects

The Ss were drawn at random from among residents of Caro State Home and Training School, Caro, Michigan, with ages from 15-40 and with IQs of 50 and greater. The total pool of Ss (N=274) was divided into two groups: the retarded (R) (N=191, IQs 50-69, $\overline{IQ}$=58.5, $\overline{CA}$=26.2) and the nonretarded (N) (N=83, IQs>70, $\overline{IQ}$=82.0, $\overline{CA}$=27.2). Group R was probably representative of similar groups selected from other institutions. Group N, however, cannot be considered "normal" in the usual sense. These Ss, in most cases, had been institutionalized just as long as the retarded Ss and remained in the institution from the time the institution served epileptics of all levels of intelligence. Persistent efforts had been made to return these nonretarded Ss to community life so that, to some extent, these remaining 83 Ss represented various behavior disorders severe enough to prevent successful community placement.

In the past, the only diagnosis of patients was of idiopathic or symptomatic epilepsy. The Home does not have, and has not had, a psychiatrist on the staff. Behavioral diagnosis has been done largely by inexperienced psychologists with bachelor's, master's or rarely, Ph.D. degrees. These diagnoses have been mainly of intelligence level and presence or absence of evidence of organicity. In general, no qualified diagnoses of behavioral disorders were available. Nevertheless, it was commonly felt among psychologists, social workers, and physicians that most of the nonretarded residents suffered from various personality disorders, emotional disturbances, or psychoses. All Ss were subject to seizures and were

taking anticonvulsants and tranquilizers at the time of testing.

A total of 84 Ss, 42 from each of the two pools, was selected at random. These Ss ranged in age from 15 to 40, $\overline{X}$=27.5 years, SD=6.67. Their IQ scores ranged from 51 to 115, $\overline{IQ}$=71.1, SD=13.9. Each S took the classification test (described below) according to which he was ranked with the other Ss to control levels of ES. The classification scores (digit spans) ranged from 2.7 to 7.5 ($\overline{X}$=4.6, SD=1.00). The characteristics of the two IQ groups are summarized in Table 9.

TABLE 9.--A summary of the ages, IQs, and digit spans of the two IQ groups

Characteristic	High IQ group (N=42)	Low IQ group (N=42)
Age[a] (years)	$\overline{CA}$=28.26 (15-40) SD=7.35	$\overline{CA}$=26.67 (18-39) SD=5.59
IQ	$\overline{IQ}$=82.69 (70-115) SD=9.84	$\overline{IQ}$=59.57 (51-69) SD=13.89
Digit span	$\overline{DS}$=5.19 (3.9-7.5) SD=.86	$\overline{DS}$=4.08 (2.7-6.5) SD=.85

[a]The two groups differed slightly in age, but this was unimportant for the present study because age is not an important variable with respect to the digit span in adults. The literature is consistent on this point.

Apparatus

A 5 x 7 photographic "safe light" (without filter) was used as a signal light to signal the time to begin reporting the digits. This lamp was operated at various times by a foot or hand switch.

Materials

Two lists of digits were prepared: one for pre-experimental classification of Ss according to their digit span and one for the experimental sessions. Both lists were constructed identically and differed simply in the specific ordering of the digits in the various strings. Each list contained 10 strings of each size from two to nine digits, plus two ascending series of strings used in pretraining. Each string was formed through random selection of digits with just one restriction: Within a string, no digit was repeated.

Procedure

The measurement of digit spans. Ss were ranked according to their performance on a digit span test given an average of 5.3 days before the experimental testing. Digits were presented ungrouped at one digit per second. There is some evidence (Blankenship, 1938) that time of day has an effect upon digit span so these effects were controlled by pretesting and by testing each S in either morning or afternoon.

The dependent variable used for the classification and experimental tests was the absolute digit span as measured by the "staircase" method. This is a psychophysical method developed for measuring absolute stimulus thresholds (Cornsweet, 1962; Underwood, 1966). The method consists of first estimating S's threshold, then constructing a series of constant stimuli around the threshold. Only these stimuli are subsequently presented. E administers an ascending series of stimuli until S detects the stimulus. As soon as S detects it, E reverses his direction and presents the next lower stimulus value. E continues to reverse direction every time

S's report changes until a constant number of trials of this type is given. S's threshold is computed by taking the mean of all stimulus values presented; the result is a value midway between detected and undetected stimulus values.

The application of this method to the determination of the digit span was straightforward. The early trials (pretraining trials) of a session consisted of an ascending series of strings, beginning with two-digit strings, through which E estimated S's digit span and through which S gained practice in the task. Two of each string size were administered in ascending order until S reported 5 strings correctly. From that point on, 1 string of each size was given until S erred on 2 consecutive string sizes. At this point S was returned to one string size smaller than his largest success and a second ascending series was given until S erred on 2 consecutive string sizes. The estimate of S's digit span was the largest string S reported correctly out of the two ascending series of strings. A string of this size was the first string presented in the series of staircase trials which began immediately following the estimation of digit span. If S reported the first string correctly, he was given a string one digit larger; if he erred on the first string, he was given a string one digit smaller. This procedure was followed for each string for a total of 9 response strings. In scoring, no regard was taken of whether S was right or wrong on a string; the final score was the mean of the string sizes presented to S. Only 9 staircase trials (strings) were given, but the score was the mean of 10 strings. This was possible because S's performance on the ninth trial determined the

size of string which would have been presented on the tenth trial had it been given. S's accuracy in reporting the tenth string was irrelevant in computing his score, so the tenth trial was not given.

The experimental conditions. A random groups design was used with levels of digit span (ES) as a control variable. The effect of grouping of digits during presentation and of variation of presentation rate were studied. The grouping variable was represented by two levels, O-grouping and grouping. Grouping consisted of reading two digits rapidly; i.e., grouping was largely temporal. The grouping was emphasized by vocal accent of the first digit of each pair and a falling intonation on the second. The O-grouping condition consisted of reading the digits in a monotone, with even temporal spacing, and speaking the final digits of each string with the falling intonation used in the grouping condition. This falling intonation conveys completion of the string and elicits prompt production of the digits.

The rate factor was represented by four levels: four digits per second, two digits per second, one digit per second, and one-half digit per second (one digit every 2 seconds in the O-grouped condition and two digits every 4 seconds in the grouped condition). The rates were controlled very simply by listening to a timer which ticked at one-quarter second intervals or by watching a stop watch. In either case, E established the rate in the few seconds before each trial then delivered the digits without close observation of either the watch or the timer.

The rate variable was crossed with the grouping variable with one cell missing. It was impractical to group digits at the rate of four digits per second because the rapid rate used to group

digits was approximately four digits per second. The use of tape recording techniques might make this combination of conditions possible, but it was felt unnecessary for this study.

The Ss of each IQ group were ranked according to the size of digit span (ES) as measured by the classification test. Starting at the top of the ranking, every seven Ss were assigned at random to each of the seven treatment combinations. As a result of this procedure, each treatment condition had one subject from each of six fairly homogeneous ES levels.

All data were collected in the cottages of the Ss. The rooms available for this purpose were removed slightly from the main areas of activity, but there were frequent noisy outbursts which interfered slightly with testing. S was brought to this room and E read the following instructions:

> I have a simple test here which will tell me how you remember numbers. The test has nothing to do with whether you stay or leave the Home; I am just interested in how people remember these numbers. The test is very easy to do and will not take long. It is just a matter of listening to me say some numbers then saying them back to me when I am done.
>
> I would like you to help me for several minutes, but you do not have to stay if you would rather not. Will you stay and help me? Thank you.
>
> Try and do the best you can, but don't worry if you forget some because some are very hard and everyone forgets them. I will read some numbers and all you have to do is say them back to me when the light goes on. Remember, don't begin saying the numbers until the light turns on. Now, let's practice some.

Results and Discussion

Classical Phenomena

The results reproduced the classic positive correlation between digit span and IQ, $r=.79$ ($t=11.68$, $df=82$, $p<.001$). This correlation is larger than the correlation found in Experiment 1 (.47

or, excluding the atypical Ss, .56), probably because the present study had a wider range of IQ scores (51-115, compared with 52.3-107.6 with only two Ss above 86.5).

The higher correlation compares well with figures reported by Jensen (1970). Jensen took correlations from the work of Terman and Merrill (1960) and Wechsler (1958) and corrected them for attenuation. So corrected, the correlation between Stanford-Binet IQs (for two and one-half year old children) and digit span was .62. Similarly, the correlation between digit span and the WAIS IQ was .75.

Since the present Ss were all adults (CA 15-40), little or no relation between age and digit span was expected. The partial correlation of age and digit span (with IQ held constant) was .14 (t=1.28, df=82, p=.10). This small correlation was of borderline significance and may reflect small changes of digit span with ages above 15 years of age. Wechsler (1958) found some growth of digit span in normal adults up to age 25. Table 10 summarizes the correlations among age, IQ, and digit span.

Practice Effects

Within-session practice. The classification test consisted of 10 test trials following a training period (4-12 digit strings depending on size of ES). S's performance on the first 5 test trials was compared to his performance on the second 5 test trials. For analysis, the Ss were divided into six ES levels on the basis of their digit spans. The data were subjected to a 2 X 6 (first half-session vs. second half-session X six levels of ES) analysis of variance for repeated measures (Winer, 1962, p. 306). The datum used was not the digit span, but was the sum of the string

sizes presented to S. In the staircase method of measurement of the digit span, the size of S's span is simply the ratio of the sum of string sizes to the number of string sizes presented in each half-session.

TABLE 10.--The correlations among age, IQ, and digit span

Variables	Correlations[a]
Digit span and IQ	.79, t=11.68 ****
Digit span and age (years)	.19, t= 1.76 **
Age (years) and IQ	.13, t= 1.19 ns
Digit span and IQ with age held constant	.79, t=11.68 ****
Digit span and age with IQ held constant	.14, t= 1.28 *
Age and IQ with digit span held constant	-.22, t=-2.04 ***

[a] df=82 in each case

ns $p>.10$

* $p<.10$

** $p<.05$

*** $p<.01$

**** $p<.001$

As hypothesized, digit spans declined significantly from the first half-session ($\overline{DS}$=4.80) to the second half-session ($\overline{DS}$=4.48) (F=11.33, df=1/78, $p<.01$). There was no evidence of an interaction of Practice X Level of ES (F=.81, df=5/78); just as in Experiment 1, Ss of all levels of ES showed a decline in performance as

a function of practice. In Experiment 1, the sessions were long; about 30 minutes were required to complete 42 digit and word strings. The present results show that the practice effect occurs over much shorter periods and with much less practice. The present classification session took about 5 to 6 minutes including pre-training (4-12 trials depending on S's digit span). The number of test strings administered beyond pretraining was 9.

The results of both studies are clear in indicating a decline in performance with practice within a session. This decline was not predicted by the S-R_e model but is not really inconsistent with it. In the discussion of Experiment 1, the decline in performance was attributed to boredom or response habituation (proactive inhibition is unlikely because of the time relations involved, see discussion in Experiment 1). The fact that a small decline occurred in this study argues against a boredom hypothesis because classification sessions required only 5-6 minutes including pretraining. Ss did not seem bored and often were surprised at the brevity of the session. The decline was probably due to response habituation.

Spitz (1966) reported a similar, but nonsignificant, effect of practice in a digit span test using simultaneous visual presentation. Retarded Ss (MA=8.6; CA=14.6) showed a decline with practice, but normal Ss (CA=8.6) did not (in contrast to the present "normal" group). Since the high IQ group in the present study was of low normal intelligence, $\overline{IQ}$=82.69, the two studies are not necessarily inconsistent.

An important result of the present analysis is the nonsignificant Practice X ES interaction (F=.81). This result means

that all Ss, regardless of size of ES, responded similarly to practice which is consistent with the sufficiency principle. The present design compensated for differences in ES and tested Ss on comparable S-R units, so it was fully expected that all levels of ES would respond similarly to experimental manipulations.

The data were further analyzed for the effects of IQ on practice. The Ss were divided into four IQ groups and the data summarized by analysis of variance. IQ did not interact with practice (F=1.07, df=3/80). All Ss, regardless of ES or IQ, showed a similar decline of performance with practice within a session.

Repeated testing. The present study also allowed examination of the effects of practice in the form of repeated testing on two different occasions. Twelve Ss (Table 11) served in the 0-grouped, one-digit-per-second condition which duplicated the conditions under which the classification scores were obtained. The effect of repeated testing on digit span performance was assessed by comparing the performance on the classification test with the performance on the experimental test. The two sessions were separated nonsystematically by from 2 to 21 days, $\bar{X}$=5.3. Two analyses were done, one with the Ss divided by digit span (the mean of both sessions) and one with the Ss divided by IQ. Both procedures produced nearly the same subject groupings (two Ss exchanged levels).

The mean digit spans for the two sessions were almost identical, 4.55 and 4.53 (F=0). Neither ES nor IQ interacted with practice (both Fs 1.0). The fact that repeated testing did not change the measured span scores is consistent with the S-R_e model which posits a constant ES ability. It also is consistent with the boredom or response habituation interpretation of the within-session

effect of practice. The effects of boredom or response habituation would obtain equally within each session but would not be expected to transfer from the first session to the second session.

TABLE 11.--The summary statistics of the 12 Ss who were tested twice under identical conditions

Characteristic	Mean	SD
IQ	75.42	19.49
Age (years)	27.00	6.63
Digit span [a]	4.54	.85

[a]The mean digit span is the average of the two sessions.

Experimental Manipulations

Rate. The effects of rate and grouping were assessed by a fixed effects analysis of variance of the digit spans as measured by the staircase method. The analysis was a 2 X 4 X 2 X 6 analysis (grouped vs. 0-grouped X one-half, one, two, and four digits per second X high IQ Ss vs. low IQ Ss X six ES levels) with ES nested within IQ. No Ss were tested under the grouped condition at four digits per second. For purposes of analysis only, these data were estimated according to a formula suggested by Winer (1962, p. 281).

Preliminary analyses of the data disclosed no evidence of a triple interaction of Grouping X Rate X ES. Since each cell of the triple interaction contained just one S, no separate error term could be calculated. The error term was dropped from the analysis and the mean square of the triple interaction served as an error term in preliminary testing.

Low IQ Ss were expected to respond minimally to the rate

variable and uniformly positively to the grouping variable. In contrast, the high IQ Ss were expected to respond uniformly and positively to rate as well as grouping: Both grouped presentation and slow presentation rates were expected to facilitate digit span performance. The data are presented in Figure 6.

As predicted, rate interacted significantly with IQ (F=5.36, df=3/58, p<.001, Table 12). Within the high IQ group, rate had a significant effect (F=7.48, df=3/38, p<.01; a separate analysis). Within the low IQ group, the rate factor was not significant (F=.86, df=3/26; a separate analysis).

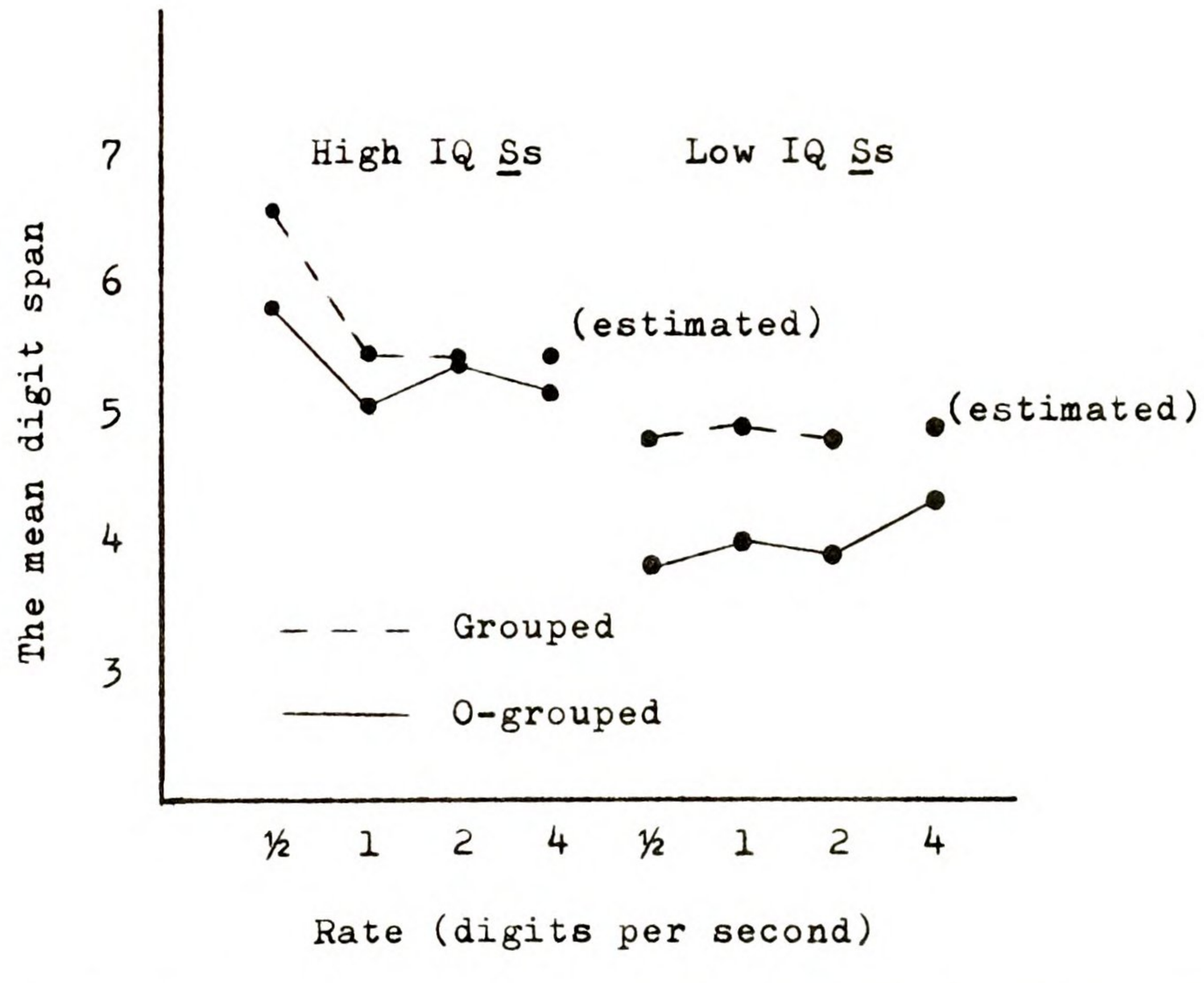

Fig. 6.--The effects of IQ, rate, and grouping on digit span.

TABLE 12.--A summary of the analysis of variance of the digit spans as a function of IQ, rate, grouping, and ES

Source	df	MS	F
Rate	3	.72	2.57 *
Grouping	1	9.44	33.71 ***
IQ	1	33.72	120.43 ***
R X G	3	.38	1.36
R X IQ	3	1.50	5.36 ***
G X IQ	1	.98	3.50 *
R X G X IQ[a]	3	.12	.43
<u>ES within the high IQ group</u>			
Levels of ES	5	6.67	23.82 ***
G X L[a]	5	.31	1.11
R X L[a]	15	.28	1.00
G X R X L[a]	15	.30	1.07
<u>ES within the low IQ group</u>			
Levels of ES	5	3.41	12.18 ***
G X L[a]	5	.22	.78
R X L	15	.47	1.68 *
G X R X L[a]	15	.23	.82

[a]Several terms in the analysis produced very small Fs so were assumed to estimate error variance and were pooled to increase the degrees of freedom of the error term. The terms were pooled according to the procedure outline by Winer (1962, pp. 202-207). Essentially, any terms which produced Fs with ps greater than .25 were pooled. The terms with the superscript a are those which were pooled. The pooled error term was .27, df=58.

* $p<.10$

*** $p<.01$

Statistically, the data were quite as expected; but within the low IQ group under the 0-grouped condition there was a slight indication of a positive relation between digit span and rate of stimulus presentation. The fact that the effect appeared only in the 0-grouped condition does not necessarily imply a Grouping X Rate interaction. The 0-grouped condition contained the four-digits-per-second rate which was not included in the grouped condition. If rate has an effect, it should show up in the 0-grouped condition which contained the widest range of rates. The fact that the digit spans of low IQ Ss in the 0-grouped condition increased as a function of rate is contrary to the present predictions, but is what Brown (1958) predicted based on a decay theory of the memory span. The mean digit span for the slowest condition was compared with the mean digit span for the fastest condition by a t for matched pairs (Ss were grouped by ES level, $t=2.38$, $df=5$). The present theoretical position (which expected no significant rate effect) implies a two-tailed test according to which t was of borderline significance ($p<.10$). Brown's position implies a one-tailed test (he predicted the result) for which t was significant ($p<.05$).

When the data for the low IQ Ss were divided according to ES level (Figure 7, below), rate appeared to have an effect opposite that on the high IQ Ss. Eight of the 12 Ss in the slowest condition had digit spans which were definitely lower than the Ss in the one-digit-per-second condition. The difference between the slowest and the next faster condition was not significant, however, (matched pairs $t=.40$, $df=11$, $p>.30$) because three Ss in the

Statistically, the data were quite as expected; but within the low IQ group under the O-grouped condition there was a slight indication of a positive relation between digit span and rate of stimulus presentation. The fact that the effect appeared only in the O-grouped condition does not necessarily imply a Grouping X Rate interaction. The O-grouped condition contained the four-digits-per-second rate which was not included in the grouped condition. If rate has an effect, it should show up in the O-grouped condition which contained the widest range of rates. The fact that the digit spans of low IQ Ss in the O-grouped condition increased as a function of rate is contrary to the present predictions, but is what Brown (1958) predicted based on a decay theory of the memory span. The mean digit span for the slowest condition was compared with the mean digit span for the fastest condition by a t for matched pairs (Ss were grouped by ES level, $t=2.38$, $df=5$). The present theoretical position (which expected no significant rate effect) implies a two-tailed test according to which t was of borderline significance ($p<.10$). Brown's position implies a one-tailed test (he predicted the result) for which t was significant ($p<.05$).

When the data for the low IQ Ss were divided according to ES level (Figure 7, below), rate appeared to have an effect opposite that on the high IQ Ss. Eight of the 12 Ss in the slowest condition had digit spans which were definitely lower than the Ss in the one-digit-per-second condition. The difference between the slowest and the next faster condition was not significant, however, (matched pairs $t=.40$, $df=11$, $p>.30$) because three Ss in the

slowest condition had unusually high digit spans for their ES levels. It is argued below in a different context that these disparate digit spans are best attributed to errors of measurement of DS. If this is the case, then it must be concluded tentatively that rate is an important variable for the low IQ Ss since 8 of the 9 remaining Ss had small digit spans under the slow condition.

According to the S-R_e model, very slow stimulus presentation rates were expected to reduce digit spans because the slow rates would allow irrelevant stimuli to interfere with performance. The rate values selected were expected to be fast enough to prevent distraction, but apparently the one-half-digit-per-second condition did allow distractions. This interpretation must be tested. One possible experimental test would involve the factorial manipulation of a distraction variable with the rate variable. The expected effect is an interaction of rate and distraction such that the effects of distractors are augmented by the slow condition. Another test would involve incentives. Theoretically, incentives should have no effect on size of ES; but they should affect S's attention to the task. An incentive condition should be most effective at slow rates where it would be expected to eliminate the rate effect observed in this study by focusing S's attention on the digit task.

The rate effect in the low IQ group is reasonably consistent with Brown's prediction from the decay hypothesis; but it can be argued that the rate effect, even if significant, is too small to support the decay theory. If the span limit is the result of loss of earlier digits through time-related decay, then doubling or

quadrupling the rate of stimulus presentation should result in a doubling or quadrupling of the span. In other words, if S can retain 5 digits in 5 seconds, he should be able to retain 10 digits in 5 seconds if they are all presented within the 5 seconds. The mean difference in digit spans between the slow condition and the fast condition was .43, but the ratio of rates was eight to one.

The literature is unclear about the effect of rate in the sequential auditory memory span paradigm. Blankenship (1938) in a review of the memory span literature to that date concluded that the data were inconclusive. Since that time, the problem of measuring the absolute digit span has not been of much interest and there are no data relevant to the rate variable in the sequential memory span paradigm. The studies which do vary rate in span-like paradigms have presented large numbers of supraspan strings and measured performance in terms of error rates of specific words or digits within a response string (for example, Mackworth, 1962; Ellis, 1970). The findings have been essentially what was found here: Normal Ss benefit by slow rates of stimulus presentation, but retardates benefit only if given special training in rehearsal techniques.

Grouping. Grouping produced a significant overall facilitation ($F=33.71$, $df=1/58$, $p<.01$) and significant effects within each IQ group ($F=7.48$, $df=1/38$, $p<.01$ in the high IQ group and $F=39.28$, $df=1/26$, $p<.01$ in the low IQ group). The Grouping X IQ interaction was of borderline significance ($F=3.50$, $df=1/58$, $p<.10$): The digit spans of the high IQ Ss were increased to a lesser extent than the digit spans of the low IQ Ss. This latter result was not

predicted but has been reported before (Spitz, 1966). The result may have been produced by differences in rates of large responses as a function of IQ groups and grouping conditions. It was concluded from the results of Experiment 1 that the digit responses of high IQ Ss are often large; that is, high IQ Ss produce large responses in an ungrouped condition. It has been assumed that the low IQ Ss do not produce large responses to an appreciable extent except when large responses are developed through practice such as echoing digit pairs in the grouped condition of the present study. The grouping operation (as a source of large responses) would have a large impact on the digit spans of the low IQ Ss because in the ungrouped condition low IQ Ss produce few, if any, large responses. The high IQ Ss produce large responses in both the grouped and O-grouped conditions so the facilitory effect of the grouping operation would not produce a strong contrast between the two conditions.

Spitz (1966) predicted and found an interaction between stimulus grouping and IQ. Spitz postulated a retardate-deficit in ability to organize stimulus information. The grouping variable is relevant to Spitz's theory because external organization of the stimulus situation might be expected to offset the organizational deficit of retardates. The grouping variable would be expected to act differentially for retarded Ss compared to normal Ss because the normal Ss are assumed to organize stimulus situations spontaneously. Spitz's experiment involved simultaneous visual presentation of digits (printed on cards). Digits were grouped into pairs by simple spacing. Grouping facilitated performance for both normal ($\overline{CA}$=8.60) and retarded ($\overline{CA}$=14.63, $\overline{IQ}$=60) Ss, but the

effect was significant only for the retarded group.

MacMillan (1970) examined the effect of grouping on digit spans of educable mentally retarded Ss (IQs=65). He presented digit strings visually by typing them on 3 x 5 cards. His grouping operation consisted of spatially separating the strings into pairs of digits and requiring Ss to read the pairs as single integers such as "twenty-seven" or "eighty-four." The main effect of grouping did not reach significance although the data were consistent with predictions. The difference between MacMillan's study and the present study may lie in the chronological ages of the Ss; MacMillan's S's were children, the present Ss, adults ($\bar{X}$=27.5 years). MacMillan's 13-year-old Ss responded to the grouping variable, while his 9-year-old Ss did not.

MacMillan interpreted his results to indicate that the grouping variable depends on developmental level. The main developmental concept of the S-R_e model is the growth of ES. It has been assumed without justification that a given ES size is equivalent theoretically without regard to age and IQ. In other words, the young child with an ES of four is dealt with theoretically just as a retardate with an ES of four. If this assumption is valid, then the present study indicated that the grouping factor is not related to developmental level in the sense of a developing capacity because the level of ES was unrelated to the grouping operation. Nevertheless, MacMillan found an age-related effect which, in terms of the S-R_e model, must be attributed to differences in the S-R pools of the two groups. In the present study the S-R pools were fairly well controlled for all Ss because simple digits are used extensively by most Ss. MacMillan, however,

grouped his digits by requiring Ss to read them as integers such as twenty-seven. His two school-age groups differed in age by an average of four years and might well have differed sufficiently in experience with compound integers to produce a developmental effect. It is interesting that the S-R_e model suggests the separation of developmental effects into an "organic" capacity variable (the ES) and an achievement variable (the nature of the S-R pool).

The facilitory effect of grouping is predicted by two theoretical points of view in addition to the present one. From the information processing point of view of Miller (1957) or Tulving & Patterson (1968), it would be expected that accented grouping of digits would encourage S to encode the digit pairs as chunks. For example, 2-3 could be encoded into twenty-three, or 1-1 could be encoded as eleven (encoding is quite different from large responses which are elicited directly). This interpretation predicts improved performance because S would recall more digits for a constant number of chunks. Spitz (1966) has interpreted the facilitation of performance due to stimulus grouping in terms of organization of the stimulus input in the sense of improving the gestalt presented to S.

The present study did not allow a test of the three interpretations of the grouping variable, but the backwards memory span may. In the backwards memory span paradigm, S simply produces the words in reverse order. Backward digit spans are smaller than forward spans (Wechsler, 1958). A chunking or encoding interpretation predicts that grouping facilitates performance in the backwards paradigm. The present interpretation of the facilitation effect does not because the present interpretation of the facilitation

effect attributes the facilitation to "echoing" of digit pairs which then occur as pairs in the response string. The backwards paradigm disrupts the order of the stimulus string so the specific digit pairs do not occur in the response string so "echoing" should not improve performance.

ES levels. It was predicted from the sufficiency principle that the level of ES does not interact with the experimental variables. Within the high IQ group, no interactions of rate or grouping with ES level were significant or approached significance (all Fs were about 1.00, analysis of variance Table 12). Within the low IQ Ss, the interaction of Grouping X ES was nonsignificant (F=1.05, df=5/26, a separate analysis); but the interaction of Rate X ES Levels was significant (F=2.24, df=15/26, p<.05, a separate analysis). In the overall analysis (the only one presented, Table 12), the pooled error term was larger which produced borderline significance (F=1.68, df=15/58, p<.10). This interaction was not expected and poses a problem for the S-R_e model.

A close examination of the data revealed that the interaction was produced by three Ss whose behavior was quite contrary to that of the other Ss. Except for the three Ss, the results are exactly as predicted. The exceptions occurred in the one-half-digit-per-second condition in level two (Figure 7). Three of these four Ss had unusually high scores compared with other Ss in their level. This result is likely due to error of measurement of DS according to which S was placed in his ES level. If this score were seriously in error, the large discrepancy could result. According to Wechsler (1958), anxiety in the test situation depresses digit

span performance which could easily explain the large scores on the second test. It is reasonable to attribute the interaction of rate and ES to error in measurement.

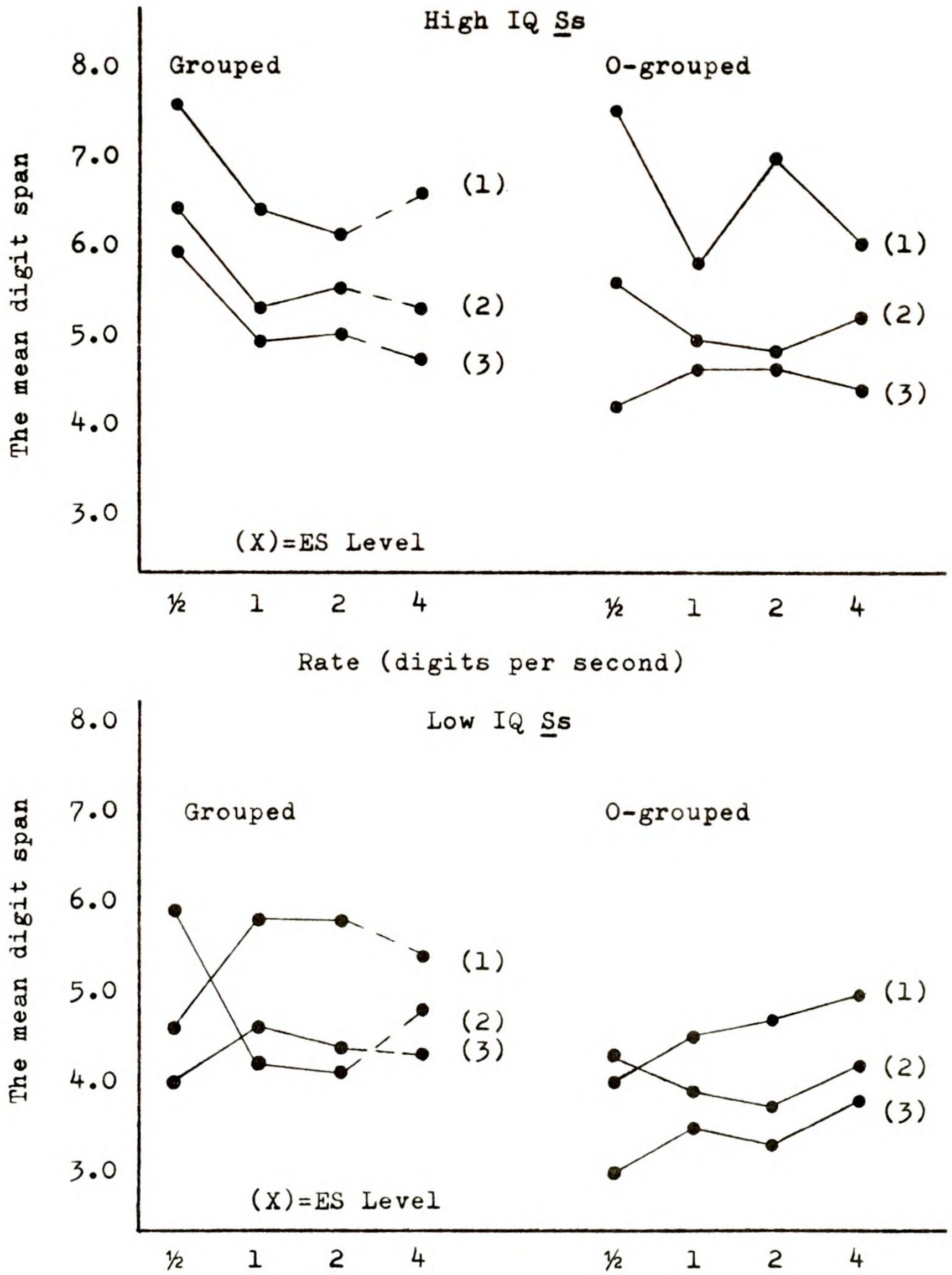

Fig. 7.--The effects of rate, grouping, and ES on digit span. The points connected by broken lines are estimated data.

Size of Effects

It is always important to judge the size of an effect to judge the importance of a variable. A highly significant variable is not necessarily an important variable in the sense of producing large differences in the dependent variable. The statistic ω^2 (Hays, 1963, p. 407) is a measure of the proportion of variance accounted for in a body of data by a given variable.

The largest source of variance was the level of ES (digit spans ranged from 2.7 to 7.5). Among high IQ Ss, ES accounted for 61% of the variance; among low IQ Ss, 42%. Ss were not selected for size of ES; they were selected at random from a normal and retarded pool of subjects. This means that the large variance produced by levels of ES is not due to selection of extreme and rare values.

When compared to the effects of experimental variables, the effects of ES were still large. The largest experimental effect was produced by grouping, yet within the high IQ group, the difference in mean digit span between the grouped and O-grouped conditions was only .52 digits. The comparable figure for the low IQ Ss was .83 digits. These differences are small compared to the range of digit spans of 3.9-7.5 in the high IQ group and the range of 2.7-6.5 in the low IQ group.

Within the low IQ group, grouping accounted for 21% of the variance, that is, one-half the variance accounted for by ES. Among the high IQ Ss, grouping accounted for less than 4% of the total variance, that is, one-fifteenth the variance that ES produced even though grouping was a highly significant variable.

The effects of rate were even weaker. Rate had no overall effect on the performance of low IQ Ss, ω^2=0, and a small effect within the high IQ group, ω^2=.10. Within the high IQ group, the average difference between the slow condition and the average of the next two conditions (one and two digits each second) was .80 digits. This effect accounted for less than one-sixth the variance accounted for by ES.

The effect of practice was highly significant but was a small effect in terms of variance accounted for, ω^2=.016. The average difference (over 84 Ss) between the mean digit spans of the first half-session and the second half-session was .32 digits (digit spans ranged from 2.7 to 7.5 digits).

General Discussion

The two experiments reproduced the classical memory span phenomena: (a) Digit span and IQ were positively correlated. (b) Word span and IQ were positively correlated. (c) Digit span and word span were positively correlated. (d) The average digit span was larger than the average word span.

New principles were suggested. The materials effect does not occur equally for all Ss but is positively correlated with IQ. In the lower IQ ranges of the first study (IQs 50-75), many Ss showed no materials effect. Performance declines steadily within a session; but repeated measurement does not affect scores, at least when test sessions are separated by two or more days. Grouping of digits during stimulus presentation increases digit spans for all

Ss (IQ 51-115) but acts more strongly among low IQ Ss ($\overline{IQ}$=59.57). Variations in rate of stimulus presentation over the range of one to four digits per second have little effect on digit spans for all Ss (IQs 51-115). At the slow rate of one-half digit per second, IQ is an important variable: High IQ Ss ($\overline{IQ}$=82.69) are likely to increase their digit spans under the slow condition, while low IQ Ss ($\overline{IQ}$=59.57) are likely to show a reduction in digit span. Grouping does not interact with rate (IQs 51-115). Individual differences in ES (size of memory span) produce considerable subject variation; but individual differences in ES do not interact with rate, grouping, within-session practice, or repeated testing.

The $S\text{-}R_e$ model. The materials effect poses a serious problem for the $S\text{-}R_e$ model. The fact that digits produce consistently higher span scores than other materials suggests that the concept of a constant ES underlying all span performance is inadequate. The present study showed that the materials effect is not a general phenomenon, but is positively related to IQ. Furthermore, the materials effect can be explained theoretically by the RSH and still retain the concept of the constant underlying ES. Tests of the RSH were supportive except in the case of transposition data. These data were not judged crucial, however, because the present design was inadequate to the task and because the theoretical status of transpositions is unclear.

In general, the results of both studies supported the sufficiency principle. This means that all Ss behave the same if the fundamental differences in ES and the specific training histories are compensated for.

In Experiment 1, practice did not interact with span level. In Experiment 2, also, practice of both types (within and between sessions) did not interact with size of digit span. Also in Experiment 2, neither rate nor grouping interacted with digit span level except in one case which was readily attributed to error of measurement of ES. Except for the interaction just mentioned, all the sums of squares for interactions produced Fs of about 1.00; in fact, the sums of squares were pooled and formed part of a pooled error term.

The most serious problem for the sufficiency principle is the fact that IQ was an important subject variable. The main effect of IQ is no problem because this simply reflects the correlation of IQ and digit span. The critical finding is that high IQ Ss responded differently to the rate variable than did the low IQ Ss. This effect is consistent with and predictable from the S-R_e model, but it would be more convincing to show that the effects of IQ can be accounted for and eliminated.

Such an experiment should not be difficult and would consist of adding the common-word condition to the design used in Experiment 2. Since common words are assumed to estimate ES directly and to be rather comparable for Ss regardless of IQ, the problem of uncontrolled large responses should be eliminated. Unfortunately, the implications of the present studies were not completely understood until the studies were completed.

Questions raised. The results of the two studies raised some problems for further analysis. In Experiment 1, the interaction between materials effect and practice was of borderline

significance. This result probably reflects a floor effect, yet such interactions within the present design are warnings either that the S-R units elicited in the experimental session were uncontrolled or another basic concept of individual differences must be added to the S-R_e model. The interaction of the materials effect and practice is particularly relevant to this point. The materials effect is defined as variation over and above that due to word span which is taken as a fairly true measure of ES. The materials effect correlated with IQ (.36) but the correlation of IQ and ES (measured by the word span) was nonsignificant (-.02 with digit span constant). It could be argued that Ss with a high materials effect (and higher IQs) have an additional ability, such as coding ability, which is related to IQ and which is independent of ES. Whereas the materials effect was effectively interpreted in terms of the characteristics of the S-R pool, more research is definitely in order.

There is some evidence that intelligence is independent of ES. It has been found that small numbers of college students have digit spans in the "retarded" range, that is, about 4.[3] This finding must be examined because of its obvious relevance to the concept of ES as a measure of intelligence. The explanation may be simple. Wechsler (1958) concluded that low spans are related to psychosis and anxiety, both of which are conceptually independent of intelligence and likely to be found at some low rate among college students.

Another problem raised is that of the definition of the

---

[3]Lester M. Hyman, personal communication, July 10, 1970.

response. The S-R_e model implies a functional definition of the response; that is, behavior which occupies one unit in ES is a response. The response has been typically defined by E in terms of consequence. An example is the bar-press. S can press the bar with any type of overt movement, all of which are called a bar-press. Another example is the word. A variety of pronunciations, either idiosyncratic or representing regional accents, are commonly accepted as instances of the same word. The classification system is based on the meaning of the word, that is, the consequence of the word for the perceiver. According to the S-R_e model, a bar-press is a response only if it acts as a response with respect to an ES paradigm. A bar-press can be smaller than a response or can consist of a sequence of responses. Similarly, a word such as catapult is probably a single response in the sense of the R in the S-R_e model. This is true, however, only for the English speaker. A foreigner speaking this word for the first time probably produces it as a sequence of phonemes, each acting as a response. One solution to the problem is to develop the concept of the ES to the point that we can be certain of the size of ES for a given S. At that point the status of a given unit of behavior might be determined by a test in an ES paradigm. If the behavior unit functions as a response and produces accurate estimates of S's true ES (determined independently), then we can conclude that the behavior unit in question is a response.

Additional tests of the ES concept must be devised. At present only the correlation of the spans produced by different materials suggests the general underlying ES. A theoretical

conception such as the S-R_e model is useful because it provides a basis for testing the sufficiency of the S-R_e model and the concept of the ES. It would be helpful to attempt to demonstrate the relevance of the ES concept to other tasks such as verbal learning or speech comprehension.

Intelligence and language. The S-R_e model has been applied only to the memory span phenomenon, but it is fairly obvious that the model is general and by implication intended for more general use. The characteristics of the ES parallel those of general intelligence, "g," which is conceived to be independent of specific behaviors.

If the IQ and the ES are compared as measures of intelligence, both measures lead to similar conclusions about variations in intelligence: (a) Both tests discriminate levels of normal and retarded behavior in a highly correlated way. (b) Both test measures grow linearly during the developmental period (Wechsler, 1958). (c) Both tests show little or no difference between males and females (Wechsler, 1958). (d) Both tests are positively related to school performance. The measures differ slightly, however: (a) IQ scores are related to race, while ES scores are less affected this way (Jensen, 1970). (b) IQ scores continue to increase slightly throughout life, but ES scores increase to age 25 then drop rather rapidly with age (Wechsler, 1958).

These two ways in which the IQ and ES differ can be reconciled if ES is taken as a measure of capacity and the IQ is taken as a measure of achievement. Wechsler (1958) pointed out that mental ability declines with age (as does the ES), but that

performance (IQ) may actually improve because of the accumulation of skills through much experience. In the case of the effect of race on IQ and ES, the IQ, having been developed and standardized on one culture, should not be expected to give exactly comparable results for quite different cultures. The ES avoids some of these problems by dealing with responses which are available and well learned for both cultures.

Although extension of the S-R_e model to a general conception of intelligence may be premature, the S-R_e model may have relevance to the concept of behavior complexity which has been theoretically related to intelligence and language.

A number of investigators have advanced a concept of behavioral complexity important in understanding several problems in learning, intelligence, and development. These concepts can be summarized simply as postulating that behaviors can be arranged into a hierarchy of complexity and that difficulty of learning is proportional to the level of complexity (Denny & Ratner, 1970; Denny, 1964; Ratner & Denny, 1964; White, 1965; Gagné, 1965).

Denny & Ratner (1970) stated that behavioral tasks differ in complexity and learning rate is inversely related to task complexity. The concept of complexity was not defined but was related to tasks involving complex cues, several of which were listed:

> (1) an extended sequence of stimuli . . . ; (2) the perseverative trace of a stimulus (stimulus after-effect) . . . ; (3) response-produced stimuli, particularly when produced by minimal responding . . . ; (4) situations in which the relevant stimuli vary with or are determined by the stimulus context; for example, when the background is light, the form of the object is relevant and when the background is dark, the color of that object is relevant . . . ; (5) situations in which the relevant stimuli must be abstracted from a larger, often changing context, as in concept formation . . . [p. 718].

The relation between these cues and phylogeny and ontogeny was summarized by Ratner & Denny (1964):

> Higher vertebrates are better able to use complex cues, as previously defined, than lower vertebrates. And, everything else being equal, more mature organisms perform complex learning tasks better than immature organisms of the same species. These trends are borne out in delayed-response learning, double alternation in a temporal maze, learning set, concept formation, and oddity learning [p. 638].

Retardates show a deficit in use of complex cues (Denny & Ratner, 1970; Denny, 1964).

> The mentally retarded rather consistently show a deficit in the area of complex learning . . . . This appears to be associated with a lessened ability to use less obvious or less available cues, as characteristic of learning set, delayed response, double alternation, oddity problem, and even problem-solving with implements [Denny, 1964, p. 120].

The first of Denny's complex cues, an extended sequence of stimuli, virtually defines the ES paradigm. The four remaining complex cues have certain similarities to the ES paradigm:

The complex cue	Similarity to ES paradigm
(1) An extended sequence of stimuli	The ES paradigm uses extended sequences of stimuli.
(2) Stimulus after-effects (delayed responding?)	S's responses in the ES paradigm are delayed during the presentation of subsequent stimuli.
(3) Response-produced stimuli	Response-produced stimuli appear against a background of external cues and increase the complexity of the total stimulus situation. Ss with large ESs respond to more aspects of a stimulus complex.
(4) Relevant stimuli which vary with the stimulus context	Ss must respond to two or more aspects of the stimulus situation to specify the correct response. Ss with large ESs respond to more aspects of a stimulus complex.
(5) Abstraction of the relevant stimulus	Ss with large ESs respond to more stimuli in stimulus complexes so they are more likely than Ss with

small spans to respond to the relevant stimulus.

These similarities between the ES paradigm and the complex cues suggest a definition of task complexity in terms of the S-R_e model; that is, the complexity of a task is the number of discrete stimuli which must be responded to in one successful trial. The memory span paradigm illustrates variations in complexity; namely, S-R is the simplest task, S_1S_2-R_1R_2 is more complex, and so on. This conception of task complexity combined with the fact that retardates have small spans predicts the Intelligence X Complexity interaction: Very few Ss indeed fail on simple S-R digit strings. In pilot work, the only Ss who failed one-word strings were profoundly retarded individuals who had almost no speech. In contrast, brighter Ss produce complex strings with ease, strings which are failed uniformly by Ss with smaller digit spans.

Not all tasks are so transparent as word strings; but they are, perhaps, measurable in terms of word string equivalents. Assuming that the principles of mixing materials are well understood, tasks of unknown complexity should be measureable in relation to a task of known complexity by requiring that Ss do both tasks simultaneously. Thus, if S's ES for words is seven, but he can, on the average, produce strings of three words correctly when he is required to add a sequence of numbers, then it could be concluded that the addition problems are equal in some sense (presumably complexity) to four-unit word strings.

Murdock (1965) did a study of this type in a test of the "limited capacity" hypothesis of immediate memory. He had Ss sort cards and recall words and found that Ss divide a limited amount of

capacity between the two tasks.

The relation between the S-R_e model and general intelligence may be pursued via a study of language (language is clearly related to intelligence; IQ tests are usually verbal tests). Denny (1966) attributed the general language deficit of retardates to the "subtle character of verbal cues [p. 5]." Verbal cues can be related to the complex cues Denny lists in another source (Denny & Ratner, 1970): (a) Sentences form long series of cues. (b) Verbal behavior requires delayed responding since S responds to a sentence usually only after hearing it all. (c) The context cue is clearly represented by the fact that the meaning of a word is determined by the context in which it occurs.

The S-R_e model is similar to language because it deals with sequences of stimuli and sequential response production. These phenomena are similar to perceiving and producing sentences. The language deficit of retardates is readily conceptualized in terms of the ES: Because retardates have small ESs, they are not so capable as normal Ss of generating the flexible novel verbal sequences which characterize language.

Language, in the sense of generation of complex novel verbal sequences, is often cited as the behavior which proves the inadequacy of "simple" S-R theorizing. Brown & Fraser (1963) comment: "Eventually children must do more than imitate and memorize if only because there is not enough time for them to learn as particular verbal responses all the sentences they will be able as adults to produce and evaluate grammatically [p. 194]."

It should be pointed out that "new" or "novel" refers to

novelty of order and sequence in S's behavior. New sentences are made of old, well practiced words, phrases, syllables, or phonemes. The S-R_e model deals specifically with this type of novelty. The novelty in S's behavior is the result of novelty in the stimulus situation. In the ES paradigm, the exchange between E and S is language in its simplest form; E presents a complex and novel (to S) sequence of verbal stimuli, and S replies likewise with a novel sequence of verbal responses closely related to the verbal behavior of E. From this point of view, the S-R_e model offers a theoretical model of intelligence, retardation, and language.

Experimental psychology is in need of a theoretically tractable conception of individual differences in intelligence. Such a conception would relate a general capacity variable to variation in learning ability at the empirical level and would do this in a theoretically obvious way. Such a capacity would vary continuously among Ss from low values to high values much as intelligence appears to do. The capacity variable could be somewhat independent of intelligence, however, reflecting the role which specific experience has in developing adaptive behavior. The capacity would grow during the developmental period as mental age does. The ES paradigm and the S-R_e model have these characteristics.

The ES Paradigm

Perhaps the most significant fact of the present studies is that Ss with IQs from 50 to 115 all served readily in the same experimental situation with no differential training. Pilot work indicated that even some Ss with unmeasurably low IQs (and digit and word spans of about one-half) perform just as readily (and

with about the same instructions) as the bright normal individual. The auditory-sequential memory span paradigm is a very compelling stimulus situation and few Ss do not respond readily.

The ease with which all Ss can serve in the ES paradigm is significant not only because of the convenience it affords E. The extensive pretraining often required to teach retardates to perform in simple experiments amounts to a confounding of pretraining and retardation. It is unlikely that the effects of pretraining are understood well enough to disregard this form of confounding.

LIST OF REFERENCES

LIST OF REFERENCES

Baumeister, A. A. Problems in comparative studies of mental retardates and normals. American Journal of Mental Deficiency, 1967, 71, 869-875.

Bijou, S. W. A functional analysis of retarded development. In N. R. Ellis (Ed.) International Review of Research in Mental Retardation, Vol. 1, New York: Academic Press, 1966.

Blankenship, A. B. Memory span: A review of the literature. Psychological Bulletin, 1938, 35, 1-25.

Brener, R. An experimental investigation of memory span. Journal of Experimental Psychology, 1940, 26, 467-482.

Bricker, W. A. Identifying and modifying behavioral deficits. American Journal of Mental Deficiency, 1970, 75, 16-21.

Brown, J. Some tests of the decay theory of immediate memory. Quarterly Journal of Experimental Psychology, 1958, 10, 12-21.

Brown, R., & Fraser, C. The acquisition of syntax. In C. N. Cofer & B. S. Musgrave (Eds.) Verbal behavior and learning: problems and processes. New York: McGraw-Hill, 1963.

Butterfield, E. C. Stimulus trace in the mentally retarded: defect or developmental lag? Journal of Abnormal Psychology, 1968, 73, 358-362.

Calhoon, S. W. A comparison of ability to reproduce one-syllable words and digits auditorially presented. Journal of Experimental Psychology, 1935, 18, 621-632.

Conrad, R. Errors of immediate memory. British Journal of Psychology, 1959, 50, 349-359.

Conrad, R. Serial order intrusions in immediate memory. British Journal of Psychology, 1960, 51, 45-48.

Cornsweet, T. N. The stair-case method in psychophysics. American Journal of Psychology, 1962, 75, 485-491.

Denny, M. R. Research in learning and performance. In H. A. Stevens & R. Heber (Eds.) Mental retardation. Chicago: University of Chicago Press, 1964.

Denny, M. R. A theoretical analysis and its application to training the mentally retarded. In N. R. Ellis (Ed.) International Review of Research in Mental Retardation, Vol.2, New York: Academic Press, 1966.

Denny, M. R. Relaxation theory and experiments. In F. R. Brush (Ed.) Aversive conditioning and learning. New York: Academic Press, 1970, in press.

Denny, M. R., & Ratner, S. C. Comparative psychology. Homewood, Ill.: Dorsey Press, 1970.

Eisenson, J. Examining for aphasia. New York: The Psychological Corp., 1954.

Ellis, N. R. The stimulus trace and behavioral inadequacy. In N. R. Ellis (Ed.) Handbook of mental deficiency. New York: McGraw-Hill, 1963.

Ellis, N. R. A behavioral research strategy in mental retardation: defense and critique. American Journal of Mental Deficiency, 1969, 73, 557-566.

Ellis, N. R. Memory processes in retardates and normals. In N. R. Ellis (Ed.) International Review of Research in Mental Retardation, Vol. 4, New York: Academic Press, 1970.

Gagné, R. M. The conditions of learning. New York: Holt, Rinehart, and Winston, 1965.

Galton, F. Supplementary notes on "prehension" in idiots. Mind, 1887, 12, 75-79.

Gundlach, R., Rothschild, D. A., & Young, P. T. A test and analysis of "set." Journal of Experimental Psychology, 1927, 10, 247-280.

Hebb, D. O. Distinctive features of learning in the higher animal. In J. F. Delafresnaye (Ed.) Brain mechanisms and learning. London and New York: Oxford University Press, 1961.

Hunt, H. F. The Hunt-Minnesota test for organic brain damage. Los Angeles: Western Psychological, 1943.

Jacobs, J. Experiments on "prehension." Mind, 1887, 12, 75-79.

Jensen, A. R. A theory of primary and secondary familial mental retardation. In N. R. Ellis (Ed.) International Review of Research in Mental Retardation, Vol. 4, New York: Academic Press, 1970.

Keppel, G., & Underwood, B. J. Proactive inhibition in short-term retention of single items. Journal of Verbal Learning and Verbal Behavior, 1962, 1, 153-161.

Kirk, S. A., McCarthy, J. J., & Kirk, W. D. Examiner's manual Illinois test of psycholinguistic abilities. Urbana: Board of Trustees of the University of Illinois, 1968.

Korst, J. W., & Irwin, O. C. Immediate memory span of mentally retarded children. The Cerebral Palsy Journal, 1968, 29(3), 10-11. (a)

Korst, J. W., & Irwin, O. C. Verbal and numerical immediate memory spans of mentally retarded children. The Cerebral Palsy Journal, 1968, 29(4), 3-5. (b)

Loess, H. Proactive inhibition in short-term memory. Journal of Verbal Learning and Verbal Behavior, 1964, 3, 362-368.

Lumley, F. H., & Calhoon, S. W. Memory span for words presented auditorially. Journal of Applied Psychology, 1934, 18, 773-784.

Mackworth, J. F. Presentation rate and immediate memory. Canadian Journal of Psychology, 1962, 16(1), 42-47.

MacMillan, D. L. Effect of input organization on recall of digits by EMR children. American Journal of Mental Deficiency, 1970, 74(5), 692-696.

Melton, A. W. Implications of short-term memory for a general theory of memory. Journal of Verbal Learning and Verbal Behavior, 1963, 2, 1-21.

Miller, G. A. The magical number seven, plus or minus two: some limits on our capacity for processing information. The Psychological Review, 1956, 63, 81-97.

Murdock, B. B., Jr. A test of the "limited capacity" hypothesis. Journal of Experimental Psychology, 1965, 69, 237-240.

Ratner, S. C., & Denny, M. R. Comparative psychology. Homewood, Ill.: Dorsey Press, 1964.

Robinson, H. B., & Robinson, N. M. The mentally retarded child. New York: McGraw-Hill, 1965.

Scott, K. G., & Scott, M. S. Research and theory in short-term memory. In N. R. Ellis (Ed.) International Review of Research in Mental Retardation, Vol. 3, New York: Academic Press, 1968.

Sidman, M., & Stoddard, L. T. Programming perception and learning for retarded children. In N. R. Ellis (Ed.) International Review of Research in Mental Retardation, Vol. 2, New York: Academic Press, 1966.

Spitz, H. H. The role of input organization in the learning and memory of mental retardates. In N. R. Ellis (Ed.) International Review of Research in Mental Retardation, Vol. 2, New York: Academic Press, 1966.

Spradlin, J. E., & Girardeau, F. L. The behavior of moderately and severely retarded persons. In N. R. Ellis (Ed.) International Review of Research in Mental Retardation, Vol. 1, New York: Academic Press, 1966.

Terman, L. M., & Merrill, M. A. Stanford-Binet intelligence scale: manual for the third revision, form L-M. Boston: Houghton-Mifflin, 1960.

Tulving, E., & Patterson, R. D. Functional units and retrieval processes in free recall. Journal of Experimental Psychology, 1968, 77, 239-248.

Underwood, B. J. Experimental psychology. New York: Appleton-Century-Crofts, 1966.

Wechsler, D. Manual for the Wechsler adult intelligence scale. New York: Psychological Corporation, 1955.

Wechsler, D. The measurement and appraisal of adult intelligence. (4th ed.) Baltimore: Williams & Wilkins, 1958.

White, S. H. Evidence for a hierarchical arrangement of learning processes. In L. P. Lipsitt & C. C. Spiker (Eds.) Advances in Child Development and Behavior, Vol. 2, New York: Academic Press, 1965.

Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill, 1962.

Zeaman, D. Learning processes of the mentally retarded. In S. F. Osler & R. E. Cooke (Eds.) The biosocial basis of mental retardation. Baltimore: The Johns Hopkins Press, 1965.

Zigler, E. Developmental versus difference theories of mental retardation and the problem of motivation. American Journal of Mental Deficiency, 1969, 73, 536-556.