Ohio Journal of Speech and Hearing[1]**,** 1977, 13(1), 5-15.

MEMORY SPAN TESTS AND THE ABILITY TO IMITATE WORD SEQUENCES AND SENTENCES[2]

Bruce L. Bachelder[3]
Western Carolina Center

The memory span test is usually thought to measure short term memory, but conceptions of the memory span test vary with the professional interests of the investigator and include learning, intelligence, information processing, and several other abilities. In an auditory-vocal span test the subject imitates the digit or word sequences of the examiner; it is hypothesized that span tests assess individual differences in the ability to imitate language. Two experiments are reported which support this hypothesis. Among retarded adolescents and adults the correlations between span and verbal imitation range from .74 to .89. One experiment found that a memory load condition reduces the ability to imitate long sentences. The implications of the experimental findings are discussed with respect to language assessment, training, and learning.

The ability to imitate the language of others is considered to be an important ability underlying the development of language skills. If a child could not imitate the speech of others he could not develop the vocabulary and grammatical constructions characteristic of his language. Language training often makes use of imitation to stimulate the acquisition of new vocabulary, phrases, and grammatical patterns (for example, Taylor, Berry, & Conn, 1976). In addition, the assessment of the ability to imitate language is frequently used as a part of language assessment (Berry, 1976; Carrow, 1974).

Since the ability to imitate language is fundamental to language development, the speech clinician will be interested in finding a way to measure this ability. When we can measure this ability directly we will be able to diagnose present and future problems in language development and prescribe efficient training techniques with the individual's imitative ability in mind.**[5]**[4]

---

[1] *Ohio Journal of Speech and Hearing* is the predecessor to *Hearsay,* published by *The Ohio Speech-Language-Hearing Association.*

[2] Reproduced with permission granted in June 2006 by Laura W. Kretschmer, Ed.D, Editor of *Hearsay*.

[3] Bachelder left Western Carolina Center in 1980 and started a full-time independent practice of professional psychology. He retired from that practice in 2003 to focus on scholarly writing and speaking. E-mail: Bruce@BruceBachelderPhD.com Web page: www.BruceBachelderPhD.com

[4] Numbers in brackets are the page numbers at the bottom of the pages of the original journal article.

The purpose of this paper is twofold: to discuss the span tests and the various interpretations of the span tests and to present research evidence for the idea that immediate memory span tests directly assess the ability to imitate language. Specific research findings will be presented which will help the speech clinician use span tests effectively in language assessment.

The Immediate Memory Span Tests

In an immediate memory span test the subject listens to a series of words or digits then tries immediately to repeat the digits or words just as they were presented. The usual measure of span is the largest number of words or digits which the subject can repeat perfectly most of the time. Span tests have long been used for assessment of intelligence, short term memory, cognitive functioning, and language. The Wechsler intelligence tests have span subtests as does the Stanford-Binet intelligence test. Span tests appear on a test for aphasia (Eisenson, 1954), on a test for brain damage (Hunt, 1943), and on a test of psycholinguistic abilities (Kirk, McCarthy, & Kirk, 1968). It has been the experience of the writer that most speech clinicians routinely use formal or informal tests of immediate memory span in language assessment.

A Broadened Concept of Span Ability

Span tests are usually considered to be measures of short term memory. For example, Williams (1965) includes span tests in a description of memory tests but not in a discussion of language assessment tests. One of the main points of this paper is that it is not necessarily true that the short term memory interpretation of span tests is the best or most useful interpretation. The psychological literature includes many quite different notions of what span tests measure.

Staats (1968) described the span test as a measure of the ability to form rote associations. Jensen (1970), also, characterized the span test as a measure of the ability to form rote associations and, therefore, as a good measure of Level I intelligence (an associative, as opposed to cognitive, intelligence). Hilgard (1951) described span tests as measures of the number of items which can be learned in one trial. Miller (1956) described the span phenomenon as one type of limitation on our ability to process information. Humpstone (1917; 1918-1919) described span ability as the ability to distribute attention over a series of discrete perceptions, and similarly, the ability to grasp a number of different elements in one operation of attention. Humpstone stressed that the span test is not a test of memory, and remarked that the conception of span as memory has marred many experiments. Leaming (1922) wrote that span tests measure the number of discrete units over which the individual can successfully **[6]** distribute his attention and still organize them into a working unit. Ellis (1963) theorized that span tests are good measures of the Hullian construct, stimulus trace. Murdock (1974) cited span phenomena as critical data to be explained by theories of short term memory. Horn (1968) cited data which indicate

that the span test is a particularly pure measure of fluid intelligence as opposed to crystallized intelligence. Fluid intelligence is considered to be the general underlying intellectual ability functioning in diverse tasks, while crystallized intelligence is considered to be the intelligence of learned responses, skills, and concepts. Bachelder & Denny (1977a, 1977b) propose that span tests measure the ability of subjects to cope with task complexity or complex stimulus control.

It is not clear which of these widely differing conceptions is most correct. The purpose of reviewing the differing conceptions is to alert the user of span tests to the widely varying interpretations of the span tests and thus broaden the application and interpretation of span tests. It is likely that the best interpretation of a span test is related to specific applications. Thus, if the span test is to be related to learning ability, then a learning conception of span ability may be the most useful. If, however, our interest is in language assessment, then perhaps yet another conception of span ability may be more useful than the short term memory conception; span tests may measure the general ability to imitate verbal sequences. That this conception is not unreasonable is obvious when one examines carefully just what happens in an auditory-vocal span test. In such a test the examiner models a word or digit sequence verbally and the subject imitates the same sequence. The following two experiments tested the validity of this conception of span ability.

EXPERIMENT I

Experiment I was done by Bachelder, Harris, Doss, & Roberto (1973). They hypothesized that if span tests measure verbal imitation ability, then span scores will correlate highly with scores on a variety of verbal imitation tasks. They measured the correlations between word span and the ability to imitate three types of word sequences: random word sequences, grammatical but nonsensical sentences, and meaningful sentences.

METHOD

Subjects

The subjects were 30 institutional adolescents (mean CA=14-1; mean IQ = 59; mean staircase word span = 3.9; mean period of institutionalization = 2.3 years). The 30 subjects were assigned to one of the three imitation conditions in such a way that the three groups of 10 subjects were approximately equal in mean span ability. This was done by ranking **[7]** the 30 subjects according to their span abilities, then the list of subjects was divided into 10 groups of three subjects each by simply marking off the first three subjects, the second three, and so on. Each group of three included subjects quite similar in span ability. The three experimental groups were formed by randomly assigning each of the three subjects to one of the three experimental conditions.

Span Test Materials

The word sequences used as stimuli to measure memory span were common single-syllable words, namely, **tree**, **hair**, **cup**, **milk**, **shirt**, **phone**, **car**, **fish**, **soap**, and **grass**. These words were assumed to be highly familiar to all subjects. It is important to minimize errors due to unfamiliarity with the words so that span scores reflect only limitations in span ability. While digits are most often used in span tests, many young normal children and retarded children and adults are not so familiar with digits as they are with the type of words used here.

Word sequences for span testing ranged in length from 2 to 10 words and were randomly generated from the pool of 10 stimulus words using a random number table. No word was repeated in a given sequence. Ten sequences of each length were recorded on Language Master cards (magnetic tape glued to card stock and played back on the Language Master Machine). The word sequences were recorded at a rate of two words per second. This high presentation rate prevents or minimizes the use of rehearsal or organizational strategies which can result in inflated span scores for higher functioning subjects (Bachelder, 1971).

Sentence Imitation Materials

Meaningful sentences for the imitation tests were constructed from common single-syllable words excluding **a**, **an**, and **the**. The nonsensical sentences were constructed from the meaningful sentences by randomly substituting other words of equivalent parts of speech for each of the words in the meaningful sentences. This procedure eliminated meaning but retained grammatical structure. The random word sequences, the nonsensical sentences, and the meaningful sentences were all generated from the same pool of words. All sequences ranged in size from 1 to 15 words. Examples of meaningful sentences and nonsensical sentences are **Large dogs chase small boys.** and **Soaps rise when at mouse.** The sentences and random word sequences were recorded on Language Master cards at a normal conversational rate.

Equipment

The materials for the span tests and the imitation tests were presented via a Bell & Howell Language Master Machine. **[8]**

Procedure--Memory Span Testing

Memory spans were measured with the staircase procedure which is based on the staircase procedure originally used to measure psychophysical thresholds (Cornsweet, 1962; Guilford, 1954; Underwood, 1966). The use of the staircase technique is a reflection of the threshold nature of memory span performance. For any given subject there is a sequence length below which performance is almost always perfect and a sequence above which performance is almost always in error. The staircase technique

provides an estimate of the sequence length half-way between these two sequence lengths. The staircase technique is consistent with Murdock’s (1974) definition of memory span as the number of items a subject can produce 50% of the time. In the staircase procedure the subject attempts several sequence lengths, some of which are just below span threshold and some of which are just above span threshold. The final score is the mean of these sequence lengths and thus provides an estimate of the sequence length which is half-way between subthreshold length and suprathreshold length.

There were three parts to the staircase span test: a) familiarization with the stimulus words, b) a set of two ascending series of word sequences and c) a set of 10 staircase trials. In the familiarization part of the test the examiner said each stimulus word singly and the subject repeated it. The familiarization procedure had three main functions: a) it established that the subject could perceive and produce each stimulus word, b) it acquainted the examiner with any idiosyncratic pronunciation (which is particularly important with retarded subjects who may have serious phonological problems), and c) it familiarized the subject with the word pool and the general test situation and procedures.

Immediately following the familiarization part of the test the subject attempted the two ascending series. This part of the test was essentially the same as the Wechsler span test. The subject first attempted a two-word sequence, a three-word sequence, and so on, until he erred on two successive sequence lengths. The examiner then began the second ascending series with a sequence length which was two words shorter than the longest sequence correctly produced in the first ascending series. The second ascending series continued until the subject erred on two successive string sizes. In the ascending series and the following staircase trials a response had to be perfectly correct with no omissions, intrusions, or order errors. The ascending series part of the test had three main functions: a) it further familiarized the subject with the test procedures (some subjects initially have trouble understanding that they are to produce all the words of a string rather than just one word), b) it provided some warm-up practice at the task, and c) it estimated the subject’s span threshold and thus determined the length of the first word sequence in the staircase trials. **[9]**

Immediately following the ascending series the examiner began the staircase trials with a sequence length equal to the largest correctly produced sequence obtained in either of the two ascending series. Whenever the subject erred, the next sequence was one word shorter; whenever he responded perfectly, the next sequence length was one word longer. This staircase procedure ensured that the subject encountered sequence lengths which closely bracketed his span threshold.

The subject’s span score was the mean of the 10 sequence lengths of the staircase trials. Note that the scoring was based upon all sequence lengths regardless of whether they

were produced correctly or incorrectly. Accuracy of performance determined the sequence lengths presented so even though accuracy was ignored during scoring it clearly determined the size of each subject's score. Since the span score was based upon sequence length, staircase trial 10 was not actually presented; the length of sequence which would have been given was determined by the accuracy of performance on trial 9. At first glance this procedure appears to give credit for performance not measured; but, in fact, each staircase sequence length reflected accuracy of performance on previous trials. The sequence length for trial 10 was determined by performance on trial 9 and the sequence length presented on trial 1 was determined by performance during the ascending series part of the test.

Before span testing began for a given subject the examiner shuffled the 10 Language Master cards for each sequence length then selected cards as needed from the top of a stack and placed the used cards on the bottom of a stack. This procedure was followed for both the ascending series trials and the staircase trials. It ensured that the selection of a specific sequence of a given length was fairly random across subjects, and that a subject did not encounter a specific sequence more than one time during a test.

Imitation Tests

The imitation test procedures were the same as the staircase span test procedures except for a slight change in scoring criteria for a given sentence or word sequence. A sentence or random word sequence had to be imitated perfectly except both intrusions of **a**, **an**, and **the** and transposition of two adjacent words were ignored in scoring.

RESULTS

Imitation spans were measured as the mean length of random word sequences or the mean number of words in the sentences during the 10 staircase trials just as in the staircase span test. Word span correlated .74 ($\mathbf{p} < .01$), .89 ($\mathbf{p} < .0005$) and .79 ($\mathbf{p} < .005$) with random word imitation span, nonsensical sentence imitation span, and meaningful sentence **[10]** imitation span, respectively. The mean imitation spans were 4.02 words, 4.14 words, and 8.54 words, for the random word sequences, the nonsensical sentences, and meaningful sentences, respectively.

DISCUSSION

The high correlations between staircase word spans and the three different imitation spans indicate that the span test does assess the subjects' general abilities to imitate widely differing types of word sequences. For this reason it appears quite reasonable to consider the span test to be a test of the ability to imitate verbal sequences whether the sequences are random words, nonsensical sentences, or meaningful sentences. Even though the imitation spans for meaningful sentences were much larger than those for the other two types of word sequences, the high correlations indicate that the relative

imitation abilities are predicted by relative span ability levels. Thus, these data indicate that the span tests can be used fruitfully to assess the general ability to imitate language, at least among retarded subjects.

EXPERIMENT II

The intent of the second experiment (Bachelder, 1975) was to replicate the findings of the first experiment and to test an implication that span ability underlies both sentence imitation and short term memory. If span ability underlies both sentence imitation and short term memory and if a subject must imitate a sentence and remember digits simultaneously, then his imitation sentence span should be reduced. This would be expected because if the subject uses some of his span ability to remember digits then he would have less span ability available to imitate sentences. In Experiment I the sentences were somewhat artificial in that they were constructed from a fairly limited pool of words and they had no articles. Experiment II used sentences which were natural language sentences**.**

Subjects

The subjects were 22 institutional adults (CA = 20.0; IQ = 56; mean staircase word span = 4.0; mean period of institutionalization = 6.8).

Materials

The materials for the staircase memory span tests were the same as Experiment I. Some examples of the sentences used in the imitation test are **Go.**, **Bring it to me.**, **I should see if I have enough money to go shopping.** and **When you eat pancakes you might like to put some butter and maple syrup on them for a sweet taste.** Seven sentences of each length from 1 to 20 words were prepared and recorded on Language **[11]** Master cards for presentation using the staircase span procedure. Seven sentences were enough so that the subject never encountered a specific sentence more than once in a test.

Procedure

Each subject took a staircase span test, a sentence imitation test, and a sentence imitation test with memory load, in that order. This order confounded practice with the experimental variable but since little or no improvement with practice occurs in the span test when the subject does not encounter identical sequences in a test (Bachelder & Denny, 1977a) this situation was judged to present little problem. In addition, any improvement with practice would run counter to the hypothesis of a memory load effect, so, if the data support the hypothesis, the possibility of a practice effect has little bearing on the conclusion. The staircase procedure was followed in each type of test just as in Experiment I. In the sentence imitation test a sentence had to be perfectly imitated; that is, unlike Experiment I intrusions and order errors were counted as errors. In the memory load sentence imitation test the subjects first heard from one to three randomly generated

digits then heard a sentence. The digits were selected from a previously prepared list of random digits selected from a random number table. No sequence was ever used twice for a given subject. The subject then attempted to imitate the sentence, then to recall the digits. If the subject imitated the sentence perfectly and recalled the digits perfectly, he was scored completely correct and on his next trial he attempted another series of digits and a sentence which was one word longer. If the subject missed the sentence but recalled the digits, it was counted as an error and on the next trial he attempted a new set of digits and a sentence one word shorter. If a subject imitated a sentence perfectly but failed perfect recall of the digits, the trial was simply discarded and he attempted another sentence of the same size with a new set of load digits. This complex scoring procedure ensured that each subject remembered the digits and imitated the sentences simultaneously. If a subject could not recall the digits, he may have simply ignored them originally in an attempt to-imitate a long sentence. The size of memory load was constant for an individual and was determined by dividing his word span by two and rounding down to a whole number of digits. Thus, if a subject had a span of 4.7, his digit load was 2 digits.

RESULTS

Sentence spans (without memory load) ranged from 3.1 to 16.1 (mean = 8.29, **N** = 22), and correlated .83 (**p** < .0005) with word span. Three **[12]** subjects were unavailable for memory load testing for reasons unrelated to the experiment and two subjects could not recall digits and imitate sentences simultaneously. The mean sentence spans with and without memory load for the remaining 17 subjects were 7.33 and 8.28, respectively. The difference between these means almost reached significance, matched **t** (16) = 1.67, estimated **p** = .06, **t** (16) = 1.746 needed for significance at the .05 level. This **t** test was unnecessarily conservative, however, because it was based upon only 17 subjects who produced measurable scores in both conditions. Nineteen subjects attempted both conditions, but two were unable to imitate sentences and recall digits simultaneously. In order to analyze the data with all 19 subjects the sign test (Siegel, 1956) was used because it requires only that the subjects be scored according to whether they performed higher, the same as, or lower than the sentence imitation condition. Since the two subjects omitted from the first analysis clearly performed lower under the memory load condition, their data could be included in a sign test analysis. Thirteen subjects performed lower under the memory load condition, five performed higher under the memory load condition, and one performed the same in both conditions. This result is significant at the .048 Level.

DISCUSSION

This experiment demonstrated again that individual differences in span ability are closely related to individual differences in the ability to imitate highly meaningful sentences. The data also support the hypothesis that span ability underlies both imitation and short term

recall because most of the subjects were able to imitate and remember simultaneously but when remembering digits their sentence spans were reduced.

The implications of this experiment extend beyond the use of the span test as a test of verbal imitation ability. The fact that a memory load reduces sentence span may be important information to the person attempting to train language skills and use. For example, a teacher might be teaching a child to use sentences to describe pictures. The teacher might show the child a picture of a boy holding a dog and present the model sentence, **The boy is holding a dog.** Assume the child imitates the sentence perfectly but that it is just within his span ability because his span is small and because he has not used this form of sentence to any great extent in the past. The teacher might present other pictures of boys holding items which the child knows and the child might well be able to continue imitating the basic sentence substituting words into the sentence according to the picture. For example, sentences as **This is a boy holding a car.** or **This is a boy holding a ball.** might be used. If, however, the teacher shows a picture of a boy holding something for which the child has no name, say, a pineapple, the teacher will have to tell him the name **[13]** which the child will then have to **remember** while attempting to imitate the model sentence. In this case the child is likely to have difficulty producing the target sentence because it requires both imitation of the model sentence and remembering the new word. In order to help the child cope with this memory load problem, the teacher can reduce the load requirements by either shortening the model sentence or by teaching the name of the unfamiliar object without the additional requirement that the child imitate the model sentence.

REFERENCES

Bachelder, B. L., The memory span paradigm: Its use for the analysis of mental retardation (Doctoral dissertation, Michigan State University, 1970) **Dissertation Abstracts International**, 32(1), 576B - 577B (1971). (University Microfilms No. 71-18, 158).

Bachelder, B. L., Sentence imitation by retardates as a function of span ability and memory load. Unpublished report 1975.

Bachelder, B. L., & Denny, M. R., A theory of intelligence: I. Span and the complexity of stimulus control. **Intelligence,** 1, 127-150 (1977a).

Bachelder, B. L., & Denny, M. R., A theory of intelligence: II. The role of span in a variety of intellectual tasks. **Intelligence,** 1(3), 237-256 (1977b) .

Bachelder, B. L., Harris, C. R., Doss, J. L., & Roberto, N. J., Span ability and imitation of random words, meaningless sentences, and meaningful sentences. Unpublished report, 1973.

Berry, P., Imitation of language by mentally handicapped children: A language assessment technique. In P. Berry (Ed.), **Language and Communication in the Mentally Handicapped.** Baltimore: University Park Press (1976).

Carrow, E., **Carrow Elicited Language Inventory: Manual.** Austin, Texas: Learning Concepts (1974).

Cornsweet, T. N., The staircase-method in psychophysics. **The American Journal of Psychology,** 75, 485-491 (1962).

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) .

Guilford, J. P., **Psychometric Methods.** New York: McGraw-Hill (1954). **[14]**

Hilgard, E. R., Methods and procedures in the study of learning. In S. S. Stevens (Ed.), Handbook of Experimental Psychology. New York: Wiley (1951).

Horn, J. L., Organization of abilities and the development of intelligence. **Psychological Review,** 75, 242-259 (1968).

Humpstone, H. J., **Some aspects of the memory span test. A study in associability.** Philadelphia: The Psychological Clinic Press (1917).

Humpstone, H. J., Memory span tests. **The Psychological Clinic: A Journal of Orthogenics for the Normal Development of Every Child,** 7, 196-200 (1918-1919).

Hunt, H. F., **The Hunt-Minnesota Test for Organic Brain Damage.** Los Angeles: Western Psychological (1943).

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).

Kirk, S. A., McCarthy, J. J., & Kirk, W. D., **Examiner's Manual: Illinois Test of Psycholinguistic Abilities.** Urbana: University of Illinois (1968).

Leaming, R. E., Tests and norms for vocational guidance at the 15-year-old performance level. **The Psychological Clinic: A Journal of Orthogenics for the Normal Development of Every Child,** 14, 193-217 (1922).

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

Murdock, B. B., **Human Memory: Theory and Data.** Potomac, Md.: Erlbaum (1974).

Siegel, S., **Nonparametric Statistics.** New York: McGraw-Hill (1956).

Staats, A. W., **Learning, Language, and Cognition.** New York: Holt, Rinehart and Winston (1968).

Taylor, J., Berry, P., & Conn, P., A study of language learning through imitation. In P. Berry (Ed.), **Language and Communication in the Mentally Handicapped.** Baltimore: University Park Press, (1976).

Underwood, B. J., **Experimental Psychology.** New York: Appleton-Century-Crofts (1966).

Wechsler, D., **The Measurement and Appraisal of Adult Intelligence** (4th Ed.). Baltimore: Williams and Wilkins (1958).

Williams, M., **Mental Testing in Clinical Practice.** New York: Pergamon (1965). **[15]**