This is to certify that the

thesis entitled

Normal Adult Performance on the Token Test

presented by

Rae B. Zimmerman

has been accepted towards fulfillment
of the requirements for

M.A. degree in Audiology and Speech Sciences

Major professor

Date May 2, 1979

NORMAL ADULT PERFORMANCE ON THE TOKEN TEST

By

Rae Zimmerman

A THESIS

Submitted to

Michigan State University

in partial fulfillment of the requirements

for the degree of

MASTERS OF ARTS

Department of Audiology & Speech Sciences

1979

ABSTRACT

NORMAL ADULT PERFORMANCE ON THE TOKEN TEST

By

Rae Zimmerman

Sixty normal, non-brain damaged adults were divided into three age groups (25 - 34 years, 50 - 59 years, 75 - 84 years). Subjects were administered Spreen and Benton versions of the Token Test and Sentence Repetition Test.

Results challenged the traditional assertion that Token Test scores are uninfluenced by age. The oldest group achieved significantly lower Token Test scores than the two younger groups. Non-significant correlations between Token Test scores and Sentence Repetition scores for each age group indicated diminished Token Test scores could not be attributed primarily to age-related changes in auditory retention.

A semantic analysis of Token Test errors was attempted.

Lower Token Test scores for the oldest age group were thought to result from their higher incidence of hearing impairment and chronic health conditions.

Information theory was employed to explain why increased Token Test error rates among older subjects might be attributed to diminishing auditory skills.

ACKNOWLEDGEMENTS

Completion of this thesis has posed something of an emotional hurdle, and as a result this paper emerged over an extended period of time. For this reason, there are numerous individuals who share collective responsibility for helping me to see its way through to completion.

The author wishes to extend sincere appreciation for the counsel, assistance, guidance and editorial skills of her committee chairman, Dr. Leo V. Deal. His advice and patience were most critical for the completion of this thesis.

Gratitude is also expressed to Linda Gillum, Ph.D. Her continued expressions of commitment, interest and enthusiasm for this project served as a needed spur throughout the prolonged writing period. Her willingness to remain a part of my committee despite her own job relocation is especially appreciated.

For Elaine Bailie, M.A., who is my friend, my mentor, and perhaps by now my weary sounding board, thanks seem so inadequate. Her continued confidence that this study would reach completion did much to ensure its finish. Her friendship and professional sharing have led me to many insights which helped to shape the thoughts finally expressed in this thesis. In addition, many thanks are due to the Bailie family who countered my frustration with perpetual support and good humor.

Many of my colleagues and friends have participated directly or indirectly in this paper. The author wishes to offer warm thanks to Jan Jones, Joan Boonin, Mary Berman, Judith Frankmann, Marlene Cosgrove, Jack A. Carroll, Michael Bailie, my co-workers in the Speech Department at Mary Free Bed Hospital, and the sixty adults who willingly took time to serve as subjects for this study. To the many unnamed others who expressed support, interest, criticism, or ideas relating to this paper, thank you.

Without the encouragement and help of Gertrude and Milton Freedland, Bea and George Schwartzmann, and Fred Freedland, this paper would never have been begun.

Finally, special thanks to Mrs. Ara Cary, for her assistance in helping me understand the implications of this thesis.

TABLE OF CONTENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER I

INTRODUCTION

The Token Test, as first described by DeRenzi and Vignolo (1962), is one of the first standardized, widely used diagnostic tests for aphasia that has been used to challenge the old notion that expressive language impairment can occur independently of receptive language impairment. Presently there is widespread agreement that expressive and receptive components co-exist (in varied proportions) within any single clinical presentation of aphasia (DeRenzi & Vignolo, 1962; Schuell, et. al., 1964; Boller & Vignolo, 1966; Eisenson, 1973; Goodglass & Blumstein, 1973). This idea has gained acceptance only recently, however, with the development of diagnostic measures of aphasic language function sensitive enough to reveal subtle breakdowns in auditory/receptive language processing.

Until it was possible to obtain firm and verifiable data, the discussion over whether expressive aphasia could occur exclusive of receptive language impairment was one which could only proceed on a theoretical, logical basis. Development and standardization of the Token Test elevated this discussion to a factual level, one where proof could be obtained through experimentation and a tentative conclusion could thereby be supported.

When the Token Test was first introduced, its authors intended it to be an improved method of assessing subtle impairments of receptive language processing (DeRenzi & Vignolo, 1962). In reviewing tests of auditory comprehension available to them at that time (e.g., the test batteries systematized and described by Weisenberg and McBride in 1935, Marie's Three Paper Test, or Head's Hand-Eye-Ear Test), they found a number of problems inherent in test construction that ultimately reduced the validity or usefulness of the tests. They concluded that a test used in determining slight receptive disorders must have very specific characteristics. First, they felt that administration time should be brief. Second, they felt excessive apparatus or materials should not be necessary, so that the test could be made readily available and easily portable. Third, they felt that stimuli should not exceed the limits of accurate auditory retention for any normal adult of any age. Fourth, they stated test item difficulty should be due to linguistic factors but should not result from intellectual or cultural factors.

DeRenzi and Vignolo went on to say:

> The difficulty must lie. . . in the lack of redundancy of the message transmitted to the patient and in the necessity, which this entails, of grasping its significance from the semantic value of every single word he hears (page 322).

They suggested several sources of redundancy that, under non-test conditions, might contribute to accurate auditory comprehension: situation cues, role cues, verbal context, and normal conversational circumlocution. They contended a test which could eliminate these sources of redundancy (linguistic and extra-linguistic), while possessing the other qualities they listed as essential for a useful diagnostic aphasia instrument, would be successful in validly and reliably assessing high level impairment of receptive language function. It was with this intention that they developed and presented the Token Test.

The Token Test Described

Although there have been a number of modifications made of the Token Test, DeRenzi and Vignolo originally conceived of it as a five part, 61 item test requiring patients to decode increasingly complex instructions given verbally by the examiner. The test was divided into five parts on the basis of item length and complexity of each stimulus; difficulty level of each part was progressively increased throughout the test.

Test items revolved around twenty tokens which were categorized by shape (circle and rectangle), size (big and little), and color (red, blue, green, white and yellow). These tokens were grouped in rows of five, each row containing all tokens of a given size and shape. For parts one, three and five, only the large tokens are used, whereas all tokens are spread before the patient for parts two and four (Figure I).

As mentioned, the test was divided into five parts. In Part I, the large rectangles and large circles are placed before the patient. Ten short commands are then given for the patient to follow. Each command requires identification of one of the ten tokens on display. The items "Pick up the yellow circle" and "pick up the red rectangle" illustrate that at this level instructions are very brief and redundancy is minimal within each item. To respond correctly, the patient must accurately decode both the object and its modifier.

Table 1: Estimated maximum number of points obtainable on the Token Test, using a weighted scoring system (DeRenzi and Vignolo, 1962). For parts I through IV, the authors disregard the verb as critical for an accurate response, as the verb is not varied.

Part	Item Example	Number of critical elements per command:	Number of items per part:	Number of elements per part of TT:
I	Pick up the yellow rectangle.	2	10	20
II	Pick up the small white circle.	3	10	30
III	Pick up the red circle and the green rectangle.	4	10	40
IV	Pick up the white large circle and the small green rectangle.	6	10	60
V	Put the red circle between the yellow rectangle and the green rectangle.	(varied) 4 to 8	21	128
TOTAL:		---	61	278

Figure 1. Array of tokens, DeRenzi and Vignolo version of Token Test. For Parts I, III and V, only large tokens (rows one and three) displayed. For Parts II and IV all tokens presented.

Part II is conducted with large and small tokens. Thus, three words must necessarily be decoded correctly in order to perform the task. Again, ten commands are given. Items typical of Part II are "Pick up the small white rectangle" and "Pick up the large blue circle."

Part III employs only large tokens but again increases the information load the patient must process. This is done by requesting him to identify two tokens at a time: "Take the red circle and the green rectangle." There are ten commands in this portion of the test.

Part IV again displays all of the tokens, arrayed in four rows before the patient. The ten commands in Part IV are similar to those issued in Part III, but the task of processing an item from the fourth section is made more difficult by the need to decode correctly two modifiers for each object in the compound sentences generated for Parts III and IV. "Take the white large circle and the small green rectangle" is an example of a stimulus the authors supply for an item in Part IV.

In the first four sections of the Token Test, DeRenzi and Vignolo simply described the type of commands the clinician should issue, and they supplied one or two examples. In Part V, however, they specifically listed the twenty-one items to be given. Each of the first four parts had good syntactic consistency; within a given part the syntactic structure remained the same, and the only differences between one item and another were found in the specific modifier(s) or shapes used in each stimulus. However, there is no such uniformity of sentence structure found in the items of Part V.

For Part V, only large tokens are displayed. Rather than merely requiring that the patient identify tokens described in successive commands, he is now instructed in various tasks he must execute using the tokens. Appendix 1 is a list of the stimulus items provided by DeRenzi and Vignolo for Part V of their version of the Token Test.

It can readily be seen, then, that performance on the Token Test is relatively free of intelligence bias, since the vocabulary for even the most difficult items is quite basic. Furthermore, it is generally assumed that the length of each command is within the limits of accurate auditory retention for normal adults, although more will be said of this later. Test difficulty arises instead from factors more inherently linguistic in nature. Steadily increasing syntactic complexity is one such factor. Increasing amounts of minimally redundant information found in successive items is a second factor and is related to the progressively increasing length of items found in consecutive parts of the test. Finally, the test is made difficult by the reduction or elimination of extra-linguistic cues arising from verbal context, situations, or clinician-patient roles.

Establishing Norms For Performance On the Token Test

The first study to examine aphasic performance on the Token Test was the one published by DeRenzi and Vignolo at the time they first presented the Token Test in the professional literature (DeRenzi & Vignolo, 1962). They selected thirteen "pure" motor aphasics and six sensory aphasics from their clinic in Milan. Each patient was previously examined and found to be free of language comprehension deficits until the Token Test was used to reassess their receptive status. They stated that patients' scores on the Token Test confirmed the presence of receptive language impairment in every instance. However, they did not reveal specific data for further discussion.

During this initial application of the Token Test to language assessment, DeRenzi and Vignolo employed a weighted scoring system for test stimuli. This weighted system would later be evaluated for its contributions to test sensitivity. The authors felt that an error on an item in Part IV, for example, should not be considered equivalent to an error in Part II.

The authors reasoned that auditory processing demands for any single item in Part IV exceeded the processing demands for an item in Part II. Compared to items at earlier levels, they felt a Part IV item was designed to contain more information requiring accurate comprehension before the command could be correctly followed. For example, correct performance on "Take the small white circle" requires correct decoding of three words: <u>small</u> <u>white</u> <u>circle</u>. For accurate execution of the instructions "Take the white large circle and the small green rectangle," six words must be understood accurately.

For this reason, DeRenzi and Vignolo decided to assign one point for every word essential for accurate comprehension and execution of the task. In the first four parts, a point was assigned to every word denoting shape, color or size. Scoring for Part V, however, also granted points to grammatic elements such as prepositions, conjunctions, adverbs, or conditionals. Part V was also the only part where the verb was not identical throughout the entire series of items at that level. While the authors did not state the total number of points possible for all 61 items of the Token Test, it is estimated that the maximum score is about 278 points for their version of the test (see Table 1).

The DeRenzi and Vignolo study was intended as a preliminary examination of the Token Test's potential for assessing receptive language impairment. While normative figures were not elicited at this time, their results supported the test as potentially useful and valid. The authors encouraged others to research their test and suggested that at some later date, when more statistical data were available, it might be worthwhile to analyze qualitatively the types of mistakes encountered most often on this test.

Four years after the Token Test was first discussed, data began to emerge in support of the test's validity. Reporting on their work with German-speaking subjects, Orgass and Poeck (1966) published results of a study performed with 66 normal subjects, 26 aphasics, and 49 patients having brain damage but no evidence of aphasia. They found that the Token Test did a very satisfactory job of disciminating between aphasic and non-aphasic performance. They determined that sex differences had no impact on scores. They also found that while patients with the highest educational levels (i.e., more than twelve years of school) made significantly fewer errors on this test than patients with the lowest level (i.e., eight years of school), no intermediate differences in performance were discernable.

Orgass and Poeck also found the Token Test to be relatively free of age bias. In examining the influence of age on the performance of normal adults having only an elementary education (8 years of school, N = 34), they concluded that age has no influence on adult Token Test scores. However, their mean age range for each of their normal age groups was reported to be 9.9 for the youngest group through 64.4 for the oldest age group. Their use of only a portion of their subjects in each age group, i.e., those having only an elementary education, in determining the relationship between age and Token Test performance might have obscured trends at the extreme age limits of their sample.

Table 2 reproduces norms for error distribution on the Token Test for a matched group of 34 normal and non-aphasic brain damaged adults subjects (Orgass & Poeck, 1966).

Table 2. Distribution of errors in normal control subjects and brain-damaged patients without aphasia. (Weighted scoring system. Table reproduced from Orgass & Poeck, 1966.)

Group	Number of errors on Token Test							
	0-1	2-3	4-5	6-7	8-9	10-11	12-13	14+
Normal subjects (N = 34)	7	9	7	5	2	4	---	---
Brain damaged subjects	5	12	4	4	5	3	---	1

One final observation came out of the Orgass and Poeck study. They found that certainly at younger ages, Token Test scores are not totally independent of age. Children under age 15 were found to make errors with greater frequency than their normal adult counterparts; this was attributed to the possibility that language skills were still developing in younger subjects.

Shortly after Orgass and Poeck's study, Boller and Vignolo (1966) released results of a study with Italian subjects. They were pursuing an answer to the question of whether "latent sensory aphasia" could always be demonstrated in patients apparently having only a reduction of expressive speech capacity.

Their findings indicated that 87% of the aphasic patients who were initially judged to have intact receptive skills showed significant reductions of their Token Test scores. They concluded that most tests typically failed to detect receptive problems but that indeed receptive language processing suffered more of an impairment than was usually noted by the examining clinician.

Additionally, Boller and Vignolo reported norms for the performance of their 31 normal subjects on the Token Test. These scores were not easily related to the norms reported by Orgass and Poeck, however, as Boller and Vignolo employed a pass/fail method of scoring items rather than using the weighted scoring system.

In 1968 Spreen and Benton published norms based on a greatly modified form of the Token Test. Employing the test as one of twenty subtests in an aphasia battery they entitled the Neurosensory Center Comprehensive Examination for Aphasia (NCCEA), they re-named their 39 item version of the Token Test as "Identification by Sentence" (IS) (Spreen & Benton, 1969b). Spreen and Benton were able to generate profiles for both aphasics and normal adults, expressed in percentiles. However, they included no data in their manual (1969b) describing sample size or age range, apparently accepting the earlier findings of Orgass and Poeck (1966) that age has no influence on Token Test scores for an adult population.

The NCCEA version of the Token Test retained the weighted scoring system initially employed for the DeRenzi and Vignolo version (Table 3). Significantly, this was also the first time performance norms on any version of the Token Test were established using an English-speaking

Table 3. Number of points available on the Token Test, using a weighted scoring system (Spreen & Benton, 1969a). For Parts A through E, the verb is disregarded as being critical for an accurate response, as the verb is not varied.

Part	Item Example	No. of critical elements per command: Color	Shape	Size	Verb	Preposition	Other	No. of elements per part
A	Show me a circle.	5	2	--	--	--	--	7
B	Show me the yellow square	4	4	--	--	--	--	8
C	Show me the small white circle.	4	4	4	--	--	--	12
D	Take the red circle and the green square	8	8	--	--	--	--	16
E	Take the large white circle and the small green square.	8	8	8	--	--	--	24
F	Touch the blue circle with the red square.	29	32	--	16	7	12	96
Total		58	58	12	16	7	12	163

population (Table 4). It is the NCCEA "Identification by Sentence" version of the Token Test that was employed in conducting the study described in subsequent chapters of this paper.

A year after Spreen and Benton published their profile for performance on the Token Test, a second study was published using the NCCEA version. In 1969 Spellacy and Spreen reported a study designed to assess performance of aphasic and non-aphasic brain-damaged patients on the Token Test. Results of their study indicated that reliability for the NCCEA Token Test was .92 with pass/fail scoring and .95 using weighted scoring. Validity was assessed as .89 for aphasic patients and .72 for non-aphasic brain-damaged patients using weighted scores.

Normative data for these two subject populations were reported by Spellacy and Spreen using both weighted scores and pass/fail scores. In discussing the advantage of employing one scoring system over another, they stated:

> The difference between the two scoring systems appears primarily to be the degree to which they differentiate among aphasic subjects. The pass-fail scoring system appears more sensitive to the range of differences within an aphasic population whereas the weighted scoring tends to remove differences within the aphasic group and accentuate the difference between aphasic and non-aphasic groups (page 396).

In the light of this analysis, it is interesting to return to the study conducted by Boller and Vignolo (1966) described above. Their study demonstrated 87% of the aphasic patients they examined who initially were judged free of receptive deficits displayed significantly reduced scores on their pass/fail version of the Token Test. If Spellacy and Spreen's interpretation of differences in pass/fail and

Table 4. Percentile rankings for normal (non-brain damaged) adults on the Spreen-Benton Token Test (1969b). Scores are obtained with a weighted scoring system.

Raw Score:	163	162	161	160	159	158	157	156	155	154	153	152
Percentile:	*	70	50	*	*	30	18	*	*	*	8	*

* Denotes no exact percentile recorded; value must be interpolated from the printed table provided.

weighted scoring systems is accurate, it is possible to speculate that Boller and Vignolo might have found an even higher percentage of their aphasic test population displayed a "latent sensory aphasia" if a weighted scoring system had been employed.

More recently, normative data describing children's performance on the Token Test have been available. Whitaker and Noll (1972) found that the test scores of young children increase with age but that the curve begins to stabilize at 7.6 of 8.0 years of age.

Gaddes and Crockett (1975) also pursued norms, specifically questioning whether the entire NCCEA battery might be used to reflect documented patterns of language development for normal 5 to 13 year old children. Like Whitaker and Noll, Gaddes and Crockett noted the greatest changes in children's performance on the Token Test between ages 6 and 8, although they continued to note changes in scores up through age 10. They noted no sex-related differences in performance among the children tested.

One final adaptation of the Token Test merits attention. In 1978, McNeil and Prescott published a commercial version of this test, which they called the Revised Token Test. Their test was actually produced as a package, including an administration manual, text book, and set of twenty plastic tokens. The Revised Token Test was carefully standardized and validated, using English speaking subjects. A 15 point multidimensional scoring system was adopted, and the test was lengthened to include ten subtests. The Revised Token Test thus represented a major departure from DeRenzi and Vignolo's original Token Test (1962) with respect to item content and method of scoring.

Statement of the Problem

While much time and attention have been focused on assessing developmental language changes in children, language changes due to aging in an adult population have been given less attention. It might be expected that such changes, if detected, would be less striking than those known to occur at the opposite end of the developmental scale. Thus, any measure of language function used in such an assessment must be sensitive enough to detect high level changes that accompany age yet retain a vocabulary that is fairly basic to all ages being examined. The Token Test, with its weighted scoring system, seems an obvious choice for an initial assessment of receptive language changes which occur with aging.

While it was the conclusion of Orgass and Poeck (1966) that Token Test scores were not affected by age, this statement was based upon an examination of 34 German-speaking subjects who ranged in age from 13 years to an undefined age above 60. The effects of age upon language performance conceivably might not have proven statistically significant, however, until age 70 or 75; in American society, a 60 year old may remain a vital and active individual for whom cognitive effects of aging remain more than a decade away.

Further undermining the conclusion that age has no influence on Token Test scores are the unpublished findings of Bailie and Lathrop (1971). They administered the DeRenzi and Vignolo (1962) version of the Token Test to twenty-four college students and found that while the mean error rate (using a pass/fail scoring method) was 2.0, the total error range was 0 to 14. Seven of their twenty-four subjects (29.1%) missed more than two items on the Token Test. The data as described by Lathrop and Bailie reflect a higher error frequency than that reported in earlier studies which reported on comparable foreign language versions of the DeRenzi and Vignolo Token Test. It is thus conceivable that youth effects as well as aging effects may influence performance on the Token Test.

While it is known that auditory retention skills show diminution with age (Spreen & Benton, 1969b), it is not known whether this change has a functional impact upon overall speech and language abilities. The Token Test, with its gradually increasing item length, might serve as a valid means of gaining insight into auditory retention and short term memory process, especially if this test is contrasted with a test which more specifically taxes auditory retention alone. Moreover, the Token Test might help to indicate whether errors begin to emerge in systematic fashion, based on word class.

Therefore, it was the purpose of this study to examine the following questions:

1. Do older, healthy non-brain damaged adults perform at significantly lower levels of accuracy on the NCCEA version of the Token Test compared to two younger adult groups of subjects?

2. Are these changes significantly related to diminished auditory retention skills?

3. Are errors on the Token Test significantly related to word type?

CHAPTER II

EXPERIMENTAL PROCEDURES

To determine the answers to the three questions which formed the basis for this study, two subtests of the Neurosensory Center Comprehensive Examination for Aphasia (NCCEA) (Spreen & Benton, 1969b) were selected for administration to 60 normal English-speaking adults, with 20 subjects tested in each of three age categories. The Token Test (TT) was the first of the two NCCEA subtests; it was selected as a subtle but general screening measure for assessing auditory comprehension. The second subtest selected from the NCCEA battery was the Sentence Repetition (SR) subtest; this was chosen as a measure of auditory retention ability.

Subjects

Participants in this study were selected for examination on the basis of age and medical history free of known cerebrovascular or neurological disease. All participants were native speakers of English. Potential subjects who demonstrated hearing or vision problems sufficient to interfere with execution of test items were eliminated in accordance with procedures described below.

Age Groups

Subjects were divided into three ten-year age brackets: 25 through 34 years, 50 through 59 years, and 75 through 84 years. In two instances, subjects were allowed to participate in the study although they technically were several months below the lower age limit for inclusion in a given age group. (Subject B.M. was 49-1 years of age but was included in the middle age group; subject D.M. was 74-7 years of age but was included in the elderly age group.) Twenty adults were tested within each age category (N = 60). Table 5 lists the age range and mean age for each of the three age groups. Table 6 lists specific data concerning age, sex, and educational level for the 60 participants in this study.

Sex

No effort was made to insure balanced subject selection for sex, as the Token Test has been reported to be free of sex bias (Orgass & Poeck, 1966; Gaddes and Crockett, 1975). Tables 5 and 6 describe the numbers of men and women who participated in this study.

Educational Level

Subjects in this study had a minimum of ten years of formal education. This was considered an adequate level of schooling, given the reported findings of Orgass and Poeck (1966) that education level significantly interacted with TT scores only for those individuals who had primary school educations (eight years or less).

Table 5. Profile of three age groups tested (N = 60), listing age range within each group, mean age, mean level of education, and number of men and women participating in the study.

Age Group:	Low 25-0 to 34-11 years	Mid 50-0 to 59-11 years	High 75-0 to 84-11 years
Age Range:	25-6 to 34-8 years	49-1 to 58-9 years	74-7 to 84-2 years
Mean Age:	29-3 years	54-3 years	78-0 years
Sex:	13 female 7 male	12 female 8 male	15 female 5 male
Mean Education Level:	12 + 4.6 years	12 + 2.5 years	12 + .85 years

Table 6. Age, sex, and education level for subjects tested within each age category.

	LOW AGE GROUP				MIDDLE AGE GROUP				HIGH AGE GROUP			
	Subject	Age	Sex	Ed. Level	Subject	Age	Sex	Ed. Level	Subject	Age	Sex	Ed. Level
1.	PZ	25-6	F	12+9	LF	54-5	M	12+4	FZ	79-5	F	12+2
2.	RR	34-0	F	12+6	MK	56-11	M	12+4	AG	76-8	F	10
3.	JJ	32-1	F	12+2	BM	49-1	F	12+3	NM	78-10	F	12+1
4.	GB	26-2	F	12+4	ME	56-11	F	12+3	GM	76-2	M	11
5.	PD	26-9	F	12	RS	53-1	F	12	HM	75-2	F	12+1
6.	MT	28-11	M	12+7	DC	50-7	F	12+3	ES	84-2	M	10
7.	CA	25-10	M	12+6	MS	55-0	F	10	AM	81-7	F	12+1
8.	SA	26-0	F	12+7	MB	54-4	F	12+4	SL	77-4	F	10
9.	MC	34-8	M	12+6	DF	51-10	M	12+12	CV	75-8	M	10
10.	SM	32-10	F	12+4	AB	50-3	F	12+2	VP	79-4	F	12+2
11.	ER	26-1	F	12+5	KS	58-0	F	12+4	SK	78-3	F	12+1
12.	BT	30-6	F	12+4	CT	54-10	M	12	MZ	76-2	F	12
13.	SA	29-2	M	12+6	FM	52-7	M	12+2	FF	77-3	F	12+4
14.	SB	30-5	F	12+3	FG	57-6	F	10	DM	74-7	F	12+1
15.	RB	29-5	M	12+5	RW	56-5	M	12+2	GM	76-7	M	12
16.	MM	26-0	F	12+3	JK	50-0	M	12+4	DD	75-1	F	12
17.	KF	33-4	F	12+6	CC	52-6	M	12	SH	77-5	F	12+4
18.	EM	26-11	M	12+1	MK	55-2	F	12+2	HG	82-1	F	12+5
19.	GL	28-9	F	12+2	IR	56-11	F	12	RQ	76-11	F	12+2
20.	WH	31-10	M	12+6	AD	58-9	F	12	RB	83-1	M	11

Medical History

Subjects selected for examination demonstrated good general health (described as freedom from known acute illness, emotional disturbance, and cognitive impairment at the time of testing). Specific questioning attempted to determine whether potential subjects had previously experienced CVA, head injury, cortical neoplasm, or other signs of neurological disease. Individuals with a positive history for any of the above problems were excluded from the test group.

Subjects with chronic diseases such as diabetes, hypertension, or renal disease were often included in this study. However, subjects having such diseases were included only if their performance during the initial interview to obtain background information proved that they were free of confusion and disorientation. Additionally, such persons were excluded from the study if they were experiencing acute symptoms associated with their illness at the time of their examination.

Justification for inclusion of subjects who reported positive medical histories for these conditions was based on a wish to maintain a representative sample of "normal" adults in each age group. Any test population containing adults who were judged entirely free of past medical problems would have been non-representative of the general population within each age group. As Weg (1975) states:

> . . . .Although there are physiological (and anatomical), psychological, and sociological changes with time, aging is not disease. Nevertheless, it is irrefutable that morbidity and mortality do go up with age. One of the major characteristic pathological changes with age is the shift from acute to chronic disease. Additional research is sorely needed to establish "norms" for levels of systemic function in middle-aged and older persons (page 245).

Appendix 2 illustrates a copy of the questionnaire employed to elicit background information from prospective subjects. Table 7 reports the numbers of subjects in each age group who reported a positive history for diabetes, hypertension, hearing loss, or other medical problems as elicited during the initial interview.

Vision

All subjects considered for examination were free of color blindness and visual field impairments which might impair performance on the Token Test. Potential subjects were questioned before inclusion in the study. Performance on initial items of the Token Test was also closely observed. If an individual demonstrated color confusion on the initial items, testing was discontinued, and all results were withheld from analysis.

Individuals normally requiring glasses or contact lenses for reading were asked to wear their glasses or lenses throughout the test. In instances where a person did not specifically require visual correction for reading, the use of glasses or contact lenses was left to his or her comfort and discretion.

Hearing

Because changes in auditory sensitivity are known to occur with age and exposure to noise (Davis & Silverman, 1970), it was not considered desirable to screen out every person showing signs of hearing loss. It was feared that rigid audiologic screening procedures might have eliminated subjects (particularly from the oldest age group) whose hearing was functional despite detectable changes in auditory sensitivity. At the same time, it was necessary to discriminate between a

Table 7. Number of participating subjects in each age group who reported a positive history for selected physical/ health conditions. Figure in parenthesis indicates percentage of age group reporting each category of impairment.

Reported History of	AGE GROUP			
	Youngest (n=20)	Middle (n=20)	Oldest (n=20)	Total (N=60)
Diabetes			4 (20%)	4 (7%)
Hypertension	1 (5%)	6 (30%)	7 (35%)	14 (23%)
Kidney Disease		1 (5%)	1 (5%)	2 (3%)
Visual field impairment			1 (5%)	1 (2%)
Hearing loss	1 (5%)	5 (25%)	7 (35%)	13 (22%)
Hearing aid			1 (5%)	1 (2%)
Seizure disorder				0
TOTAL NUMBER OF REPORTED PROBLEMS:	2	12	21	35

subject who made an error because of hearing loss and a subject whose error resulted from impaired auditory retention or auditory processing ability.

Participants' hearing adequacy was assessed within the context of the on-going test situation. Subjects', performances on the Sentence Repetition (SR) subtest of the NCCEA battery were inspected for frequency of error. Individuals whose error rate fell below the twenty-fifth percentile on the normal adult profiles for this test (Spreen & Benton, 1969b, page 8) were asked to undergo pure tone audiometric screening to rule out the presence of hearing loss as a source of error on the tests. In instances where formal auditory screening was conducted individuals were retained as subjects if pure tone hearing thresholds fell within 10 dB of the hearing thresholds established as normal for a person of his or her age, according to the hearing threshold/age profiles reported in Davis and Silverman (1970, page 111).

Following this procedure, only 6 subjects required pure tone screening. Hearing thresholds were within normal limits for each individual (Table 7).

Individuals who wore a hearing aid for a known hearing loss were permitted to participate in this study. Again, performance on the Sentence Repetition subtest of the NCCEA was used as a criterion for inclusion in the test group. People were considered suitable candidates if their performance (with hearing aid) on the SR ranked above the twenty-fifth percentile for normal adults. Only one of the participants in this study (E.S. in the high age group) required a hearing aid.

Test Materials

Subjects were given the Spreen-Benton version of the Token Test (TT) (1969b), which is otherwise called "Identification by Sentence" (subtest 11, NCCEA battery), and the "Sentence Repetition" (SR) subtest taken from this same battery.

There were several reasons why the Spreen-Benton version of the Token Test was considered better suited to this study than versions discussed earlier in Chapter I. First, the Spreen-Benton TT was the one version used in studies with English speaking subjects when norms were derived. (The Revised Token Test by McNeil and Prescott, 1978, was not commerically available when this study was initiated, although it, too, was standardized with English speaking subjects.) Second, the Spreen-Benton TT employs a weighted scoring system which is more appropriate for error categorization than simple pass/fail scoring (Spellacy & Spreen, 1969). Third, a percentile profile for the performance of normal adults had already been published for this version of the Token Test (i.e., for all NCCEA subtests), although there is no breakdown available for data regarding age or error type. Finally, it was considered an advantage to employ two subtests from the same aphasia test battery. Given the use of the SR as a measure of auditory retention in this study, it was felt that the NCCEA version of the TT would allow a direct comparison between two subtests that were standardized on the same reference population.

Token Test

Appendix 3 includes a list of the 39 commands which compose the NCCEA TT. Administration and scoring was in accordance with guidelines expressed in the NCCEA manual, except that administration of succeeding items was not discontinued after three successive failures. It was reasoned that the purpose of the study was to examine performance of normal subjects on the Token Test rather than to pro-rate scores on the basis of inferred performance in instances where subjects showed relatively high error frequencies.

A modified answer sheet, adapted from the NCCEA, was used to record errors (Appendix 3). Each subject's exact response was noted at the time of administration. Errors were categorized with respect to semantic word type. Categories included color, shape, size, verb, preposition, and other. (Errors falling into the other category include mistakes due to incorrect recognition of adverbs, conjunctions, negation, conditionals, or the plural /s/ morpheme.)

Sentence Repetition

Appendix 4 identifies the 22 phrases and sentences which compose this subtest of the NCCEA. Subjects were instructed to repeat each sentence as accurately as possible, using the introduction to this subtest which is described in the NCCEA manual.

Sentence stimuli for the SR subtest were recorded on a Memorex MRX_2 Oxide cassette tape, using a Superscope CD-302 cassette deck for recording purposes. The speaker was a female who spoke Standard English (General American). On the SR tape, the items were preceded by an introductory paragraph and two sample items. This allowed the participating subject to set the volume control at an adequate level for accurate and comfortable listening.

The introductory paragraph on the SR tape was the following:

> Now you have heard the instructions for this test. You are about to hear two trial sentences. Set the volume control at a level that is comfortable for accurate listening, and repeat each trial sentence after you hear it. (Pause)
>
> Please sit down. (Pause for response)
>
> Cotton grows in warm countries. (Pause for response)

Ten seconds after the introductory paragraph and two sample items were presented on the cassette tape, the SR items were recorded. Each stimulus item was separated from subsequent items by approximately a five second interval. Subjects were given unlimited time, however, in which to respond. Self-corrections were permitted, but in no instance was a stimulus item repeated.

A Sony TC-110A portable cassette tape recorder with remote control switch or a Craig 2603 portable cassette tape recorder with a remote control switch was used to present the SR stimuli. Loudness levels were set by the examiner at a comfortable listening level judged sufficiently loud for accurate reception of test items. However, subjects were encouraged to modify the volume setting during the introductory phase of the SR should they find it desirable. In most instances no adjustment was necessary.

Items were presented in a free sound field, in a manner consistent with the standards presented in the test manual. Testing was done in a quiet room free of visual or auditory distraction.

Method of Presentation

To prevent the results from reflecting a bias because of order of test presentation, one-half of the subjects within each age group were given the TT followed by the SR test. The other half of each age group received the SR first. A three minute interval between the TT and SR minimized any effects of auditory fatigue.

Prior to initiation of the experimental tasks, all subjects were interviewed to assure their conformity with selection criteria outlined earlier in this chapter.

CHAPTER III

RESULTS AND DISCUSSION

Question I: Relationship Between Age and Token Test Performance

Table 8 displays the list of Token Test (TT) raw scores achieved by the 20 participants in each age group. For the young age group (25 through 34 years), a mean score of 162 out of a possible score of 163 was obtained. Individual TT scores for this age group fell within a 4 point range. The mean score for subjects in the middle age group (50 through 59 years) was 161.2, with a range of 9 points between the lowest and highest scores in that group. For the oldest age group (75 through 84 years), a group mean of 158.6 was obtained on the Token Test. For this group, there was a 21 point range between lowest and highest scores. Figure 2 is a frequency distribution for TT scores for each age group.

The data contained in Table 8 were subjected to a one-way analysis of variance according to methods described by Williams (1968). As a general formula, Williams defines F as the variance between groups divided by the variance within groups. Employing mean squares (MS) as a measure of variance, a value of 7.95 was computed for F. This was found to be significant at the .01 level (Table 9).

A Tukey-Snidecor test was then executed to determine where the significance between age groups arose (Linton & Gallo, 1975). The

Table 8. Token Test scores. (Maximum correct score is 163.)

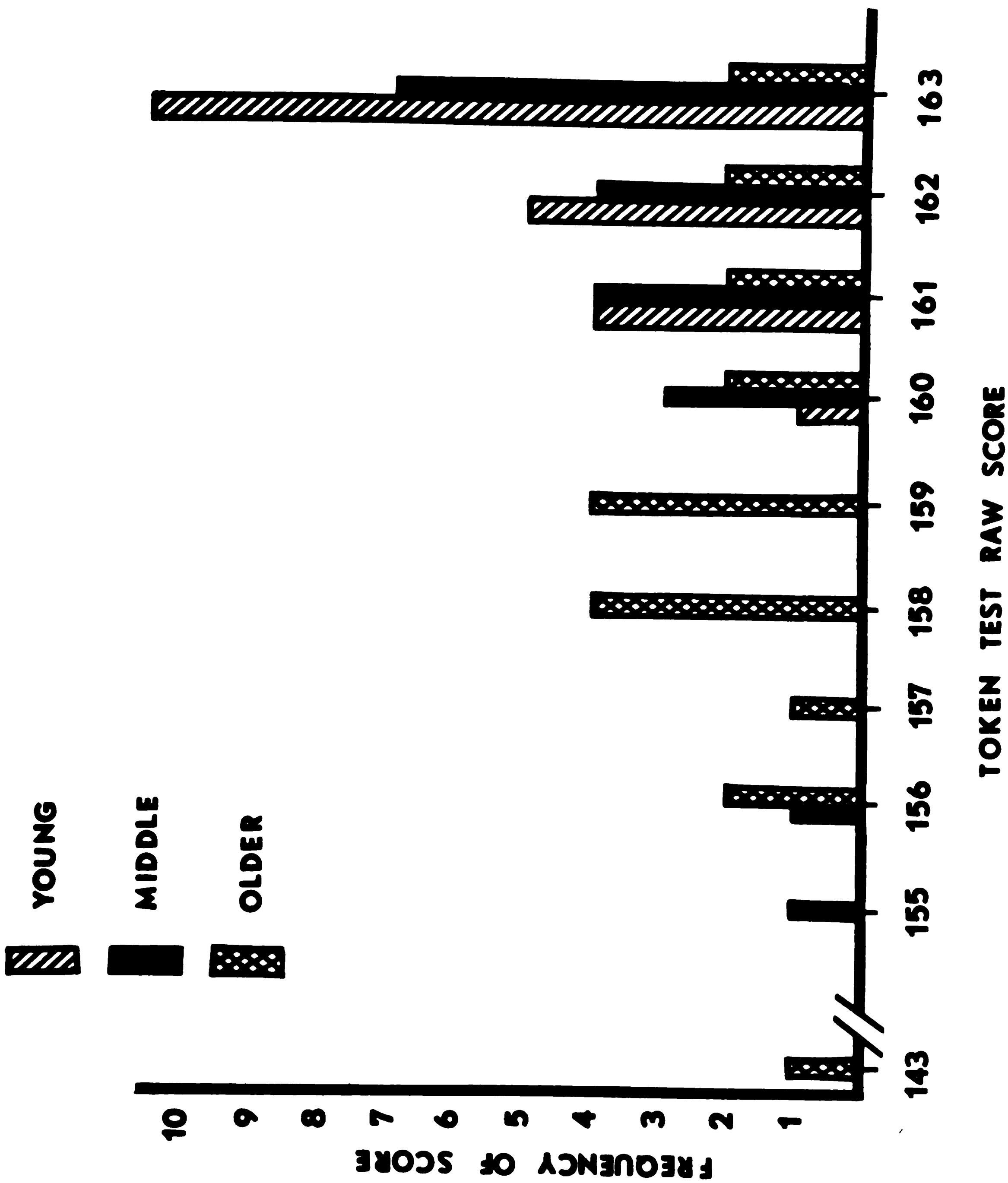
	Young	Middle	Old
Subject	Raw Score (correct)	Raw Score (correct)	Raw Score (correct)
1	163	161	156
2	162	160	156
3	162	161	158
4	163	163	157
5	163	163	158
6	162	160	160
7	161	162	163
8	162	163	159
9	163	162	160
10	162	162	143
11	163	163	162
12	163	162	159
13	163	163	162
14	163	161	158
15	161	163	159
16	161	163	161
17	163	156	161
18	160	161	163
19	161	160	158
20	163	155	159
TOTAL:	3244	3224	3172
MEAN:	162.0	161.2	158.6

Table 9. Summary, one way analysis of variance.

Source	SS	d.f.	MS	F
Between	126.60	2	63.3	7.95*
Within	453.44	57	7.96	
TOTAL:	580.04			

*Significant at the .01 level

Figure 2. Token Test Frequency Distribution For Raw Scores


YOUNG
MIDDLE
OLDER
FREQUENCY OF SCORE
10
9
8
7
6
5
4
3
2
1
143
155
156
157
158
159
160
161
162
163
TOKEN TEST RAW SCORE

significance level chosen to support the hypothesis was .05. This test requires paired comparisons of three age groups. The difference between groups is considered significant if the difference between the two mean group scores undergoing comparison is equal to or greater than the critical value calculated for a pre-set confidence level. Table 10 is a summary of this procedure.

Table 10. Summary of Tukey-Snidecor Test of Significance

Age Group	I	II	III
Mean TT score	162	161.2	158.6
Group I	---	0.8	3.4**
II	---	---	2.6*
III	---	---	---

**Significant at .05 and .01 levels; critical value = 2.696
*Significant at .05 level; critical value = 2.142

As may be seen, significant comparisons at the .05 level exist between the youngest and oldest age groups and between the middle and oldest age groups. There is no significance to the differences between the young and middle age groups. This confirms the fact that the subjects in the oldest age group performed significantly lower on the Token Test than younger individuals.

Question II: Are Age-Related Differences in Token Test Performance Related to Diminished Auditory Retention?

Subjects were given the NCCEA version of the Token Test (Spreen & Benton, 1969b) and the Sentence Repetition subtest (SR). The SR subtest was considered a measure of the ability for a subject to retain or recall auditory stimuli. Henceforth, it will be referred to as a measure of auditory retention ability. Thus, to determine whether diminishing auditory retention skills were associated with the age-related changes in TT performance described in Question 1, Pearson r correlations were calculated for each age group and for the total population (N = 60). Table 11 lists the raw TT and SR scores for each subject; the products of these scores were used in calculating r according to procedures described by Ebel (1972).

For the young age group (25 through 34 years), r = .283. The middle age group (ages 50 through 59 years) had a somewhat higher correlation between TT performance and auditory retention as measured by the SR. For the middle age group, r = .404. The oldest age group (75 through 84 years) showed a correlation of .365 between TT performance and SR performance. Thus, the oldest age group obtained the median correlation rank.

While Spreen and Benton (1969b) offered a method of applying an age correction factor to raw scores on the SR for people over 35 years of age, use of the correction factor did not influence r values. Thus, mention is made of the Pearson r for raw data, although these numbers apply to age-corrected scores as well.

Table 11. Raw Token Test and Sentence Repetition scores used in calculating Pearson r. (Maximum Token Test score is 163; maximum Sentence Repetition score is 22.)

	Young		Middle		Older	
Subject	TT	SR	TT	SR	TT	SR
1	163	18	161	19	156	14
2	162	18	160	16	156	13
3	162	16	161	14	158	14
4	163	15	163	16	157	14
5	163	19	163	18	158	18
6	162	14	160	13	160	11
7	161	16	162	14	163	15
8	162	17	163	15	159	12
9	163	16	162	14	160	16
10	162	15	162	19	143	12
11	163	15	163	19	162	12
12	163	15	162	14	159	14
13	163	19	163	12	162	18
14	163	16	161	17	158	15
15	161	16	163	18	159	15
16	161	15	163	14	161	16
17	163	15	156	14	161	16
18	160	16	161	14	163	16
19	161	14	160	12	158	13
20	163	17	155	12	159	12
Total:	3244	322	3224	304	3172	286

When data for all 60 subjects were combined, the correlation between TT and SR raw scores was strengthened. Pearson r was .440 for the total population. Table 12 summarizes these results.

Question III: Are Token Test Errors Related to Word Type?

Because adjectives, verbs, adverbs, prepositions, and other categories of "critical elements" in the Token Test occur with unequal frequency (Table 3), it was difficult to compare directly the error frequencies for difference word types. Moreover, word position within each sentence, rather than word type, might in some fashion be related to error frequency. Nevertheless, it was considered desirable to make a general comparison of error types between the three age groups to determine whether any discernable pattern might emerge.

Critical elements on the Token Test were divided into the following categories: color, shape, size, verb, preposition, and other. While other ways of categorizing these elements were considered, it was desirable to employ word categories that were not too general category in classifying TT errors, for example, all color and size errors (i.e., 70 possible errors out of 163, or 42.9% of the elements on the TT) would fall into a single class. On the other hand, a more specific word class might have been designated which would have too small a frequency of occurrence in the TT to permit reliable comparisons between groups. For example, adverbs might have been designated as a word class for purposes of error analysis, but only two elements out of 163 are adverbs in this version of the Token Test.

Table 12. Correlations between Token Test Raw Scores and Sentence Repetition Raw Scores.

Age Group	n	correlation (r)
Young	20	.283
Middle	20	.404
Older	20	.365
Overall	60	.440*

*Significant at the .01 level.

The category of "Other" was used to tally any TT error which did not properly fall under the categories named above (color, shape, size, verb, preposition). Errors in this category might include failure to acknowledge plural /s/ morphemes, as in item 34 ("Touch the squares slowly and the circles quickly"). Mistakes resulting from failure to honor adverbs, conjunctions, negation, or conditional phrases were also recorded in this category.

A "percentage of error" for each word class was calculated for individual age groups as well as for the total subject population. This was done by taking the number of errors committed within a given word class and dividing this by the total number of errors committed in all word categories by subjects in that age group. The resulting percentages were compared against the other two age groups, and against the composite group, N = 60 (Table 13).

Table 13 shows the error distribution for the young age group. A total of 16 errors were made by the 20 subjects in this age group. Prepositions accounted for 31.3% of the committed errors, whereas 25% of the errors resulted from incorrect shape identification and 25% of the errors resulted from faulty color identification. No errors relating to size or verb were recorded. Of the total number of errors committed 18.8% were classified as 'other.'

Table 13 also shows the error distribution for the middle age group. This group made more errors (36) in more categories than the young age group. For this group, the largest error category was 'other,' accounting for 44.4% of the mistakes. Shape errors were the second most frequently occurring type of error, amounting to 19.4% of the incorrect responses made by this age group on the TT. Incorrect

Table 13. Distribution of errors, Token Test.

Age Group	Errors	Error Type: Color	Shape	Size	Verb	Preposition	Other	Total Errors
Young	no.	4	4	0	0	5	3	16
	%	25%	25%	--	--	31.3%	18.8%	100.1%
Middle	no.	6	7	3	0	4	16	36
	%	16.7%	19.4%	8.3%	--	11.1%	44.4%	99.9%
Older	no.	14	16	3	1	21	33	88
	%	15.9%	18.2%	3.4%	1.1%	23.9%	37.5%	100%
Total	no.	24	27	6	1	30	52	140
	%	17.1%	19.3%	4.3%	0.1%	21.4%	37.1%	99.3%

color identification accounted for 16.7% of the errors. Only 11.1% of the errors by the middle age group resulted from incorrect preposition comprehension, as compared to the young age group where this was the most frequently occurring error type. Finally, 8.3% of the TT errors resulted from shape confusions with this age group.

The oldest age group made a total of 88 errors on the TT (Table 13). Of the errors produced by this group, 37.5% fell into the 'other' category; error frequency for this category was much closer to that incurred by the middle age group (16 errors, 44.4% of the total number of errors committed by the middle age group) than to the frequency of errors committed by the young age group (3 errors, 18.8% of the total number of errors committed by the young age group). For both oldest and middle age groups, the 'other' category was the highest ranking source of error, although it was the lowest ranking source of error for the young age group (Tables 14 and 15).

Of the TT errors by the oldest age group, 23.9% fell into the preposition category. Shape errors accounted for 18.2% of the errors within this age group. Color confusions accounted for 15.9% of the errors. Errors in the size category contributed to 3.4% of the total errors committed. Verbs were the least frequently occurring category of error, amounting to 1.1% of the errors produced by the 20 subjects in the oldest age group.

A composite frequency distribution for all TT errors made by all 60 subjects was obtained by adding the group totals for each word category (Table 13). Overall, 37.1% of the errors occurred in the 'other' category. The next most frequent error type was the preposition category, with a 21.4% occurrence rate. Shape accounted for 19.3% of

Table 14. Frequency of errors within Token Test word-class categories. Arranged in order of decreasing occurrence. (% = number of errors in word category divided by total number of errors committed by subjects in group.)

Young		Middle		Oldest		Composite	
Word Class	%	Word Class	%	Word Class	%	Word Class	%
Prep.	31.3%	Other	44.4%	Other	37.5%	Other	37.1%
Color	25.0%	Shape	19.4%	Prep.	23.9%	Prep.	21.4%
Shape	25.0%	Color	16.7%	Shape	18.2%	Shape	19.3%
Other	18.8%	Prep.	11.1%	Color	15.9%	Color	17.1%
---	---	Size	8.3%	Size	3.4%	Size	4.3%
---	---	---	---	Verb	1.1%	Verb	0.1%

Table 15. Ranking of word classes on Token Test re: frequency of error.

	Rank:			
Category	Composite	Young	Middle	Oldest
Other	1	4	1	1
Prep.	2	1	4	2
Shape	3	2.5	2	3
Color	4	2.5	3	4
Size	5	-	5	5
Verb	6	-	-	6

the TT errors, closely followed by color which represented 17.1% of the errors produced on the TT. Incorrect identification of size resulted in 4.3% of the errors; 0.1% of the total number of errors resulted from incorrect verb recognition.

In comparing error frequencies between age groups to the error frequencies for the composite group (N = 60), it must be kept in mind that the higher error rate of the oldest age group will greatly influence the percentages reported in Table 13, as this group of 20 individuals produced 62.9% of the errors recorded for all 60 subjects participating in this study.

DISCUSSION

The Token Test and Age Group Performance: Discrepancy in Research Outcomes

Administration of the Token Test to the three different age groups yielded an outcome in contrast with reported findings by previous researchers (Orgass & Poeck, 1966). In one of the first studies to emerge following publication of the Token Test by DeRenzi and Vignolo in 1962, Orgass and Poeck found that Token Test scores were age-independent for adults, although there was a higher frequency of errors when subjects ranged below 15.0 years of age. They wrote:

> . . .Although it is somewhat surprising that age has practically no influence on Token Test scores, this finding is not entirely unlikely. It is in agreement with the common experience that language performance is relatively resistant against the general intellectual deterioration of old age (page 240).

This contention, that adult performance on the Token Test is unbiased by age, has been accepted quite widely over the past fifteen years (Spellacy & Spreen, 1969; Swisher & Sarno, 1969; Van Dongen & Van Harskamp, 1972). Moreover, Spreen and Benton (1969b) state in their NCCEA Manual that Token Test scores do not require age correction computations, unlike auditory retention measures such as their Sentence Repetition subtest. This implies that Token Test performance is considered free of age effects by these investigators as well. What, then, may account for the discrepancy between data reported by Orgass and Poeck (1966) and the data found in the present study?

A partial explanation of the difference in findings may be obtained by close examination of the population of subjects whose Token Test scores were considered in the Orgass and Poeck study. They tested 141 German-speaking subjects, a group which included 66 normals, 49 patients with known lesions of the peripheral nervous system but without aphasia, and 26 diagnosed aphasic patients. Table 16 lists the number of normal (non-brain damaged) subjects who were tested within each of their reported age groups.

Table 16. Distribution of subjects within age groups, adapted from Orgass and Poeck (1966).

Age range	All normals in age category	Normal Ss with elementary education
5 - 14 years	15 Ss	(disregarded)
15 - 29 years	15 Ss	7 Ss
30 - 44 years	12 Ss	12 Ss
45 - 50 years	12 Ss	7 Ss
60 & over	12 Ss	5 Ss
Total:	66 Ss	31 Ss

While a total of 66 normal subjects were tested, only 31 of these subjects' Token Test scores were utilized in calculating their age/performance analysis of variance. Subjects below age 15 were not included in their computations, and those subjects who had more than an elementary education (i.e., more than 8 years of school) were also not included in their calculations.

Inspecting the smaller test population that Orgass and Poeck actually studied in arriving at their conclusion that Token Test performance is independent of age, it becomes more evident that they were working with a small number of subjects (five) beyond age 60. They did not report a ceiling age for their 60-and-over age group; it is conceivable that their oldest subjects were clustered around the 60 to 65 age bracket and thus were more than a decade younger than the mean age of the 20 oldest subjects participating in the present study. That is, their failure to obtain significant differences between their oldest and youngest age groups might be a function of the relatively 'youthful' ages of the five subjects who were included in their senior age group.

The outcome of the Tukey-Snidecor test provides further evidence to support the conclusion that the two studies may differ in outcome as a result of the differences in age groups selected in the two respective studies. As described earlier in the Results section, F obtained significance when the oldest age group was compared to either of the younger two groups, although no significant differences were shown to exist between the young and middle age groups in that comparison (Table 10). This means that the trend toward diminished Token Test performance is not significant until some point beyond the age level bracketed by the middle age group (50 to 59 years). The Tukey-Snidecor test shows that the significance arises during some interval beyond age 59 and that by the time a mean age of 78-0 is reached, group performance on the Token Test is already showing a significant decrement in test scores.

For their research design to be sensitive to the changes in Token Test performance that might arise with aging, Orgass and Poeck would have needed to extend their age categories well beyond 60 and would have needed to increase the number of subjects studied in the higher age category.

There is some consistency between outcomes of the current study and the one conducted by Orgass and Poeck if we are willing to adopt a supposition. The supposition is that the oldest age group selected by Orgass and Poeck probably had a mean age group, were the data available, that is closer to the middle age group employed in the present study (i.e., those in the 50 through 59 year age category). If this hypothesis is accepted, then both studies would be found to conform in the finding that there is no evidence of significant differences in TT performance for adults of any age category below the approximate age of 60 years.

A second reason, also procedural, that might account for the different conclusions regarding significance of age on Token Test performance arises in considering the differences in educational backgrounds between subjects participating in the Orgass and Poeck study and in the present study. Orgass and Poeck utilized Token Test data from only those subjects who had an elementary education (which they defined as eight years of public school). However, the current study rejected all subjects who had less than 10 years of school. Mean education levels of all three age groups were beyond 12.0 years (Table 5). It is possible, then, that discrepant conclusions regarding Token Test performance and aging effects reflect the smaller sample size of the ORgass and Poeck study, the possible differences in age ranges examined by the two studies, and the educational/cultural differences between the two test populations.

The Token Test and Auditory Retention

Thus far, the present data indicate that Token Test scores show significant but slight reduction with increased age of normal listeners. It also indicates that this age influence is evident beyond the 50 - 59 year age interval, which is the age level beyond which statistical significance between age groups was demonstrated. As we are discussing normal, non-brain damaged adults, some effort must be made to advance an explanation which accounts for this reduction in test performance with age.

One possible explanation was projected and incorporated in the design of the current study. Superficial observation of Token Test items (Appendix 3) shows an elementary conclusion: items increase in length as successive parts of the test are administered. Therefore, an effort was made to obtain a correlation between Token Test performance and performance on the Sentence Repetition subtest from the NCCEA. The Sentence Repetition subtest was intended to be a measure of auditory retention ability. It was reasoned that auditory retention, thought to be a declining function with increasing age (Spreen & Benton, 1969b; Lezak, 1976), might account in part for the significant differences in Token Test performance between young and old subjects.

As revealed in Table 12, correlations between TT and SR scores were fairly low. When a test of significance for r was applied, the correlation for the complete subject population (N = 60; r = .440) was found to be significant at the .01 confidence level. While this would appear to support the fact that auditory retention ability has some general interaction with TT performance, findings for the individual test groups at every age level failed to obtain significance for r at either the .01 or .05 levels. It would seem, then, that any attempt to explain why TT scores tend to decline with age must not rely too heavily on superficial reasoning that auditory retention skills, which decline with age, tend to render the last sections of the TT more difficult for older subjects, in turn inducing a somewhat higher rate of error.

Correlates of the Aging Process

Age-related changes in auditory retention ability are insufficient to account for differences between young and old age performance on the Token Test. A consideration of some of the medical or physiological differences between the three age groups might help to generate a more competent theory explaining the differences in TT performance which arise with advancing age.

As a person grows older, his or her chances of experiencing diabetes, heart disease, arteriosclerosis, hearing loss and numerous other conditions progressively increase. While these problems may be insufficient to interfere with normal day-to-day activity, they may perhaps induce subtle alterations in an individual's ability to function at former optimal levels. For example, arteriosclerosis and diabetes may result in small vessel occlusion in the cerebral cortex, although neurologic findings might be non-specific. Heart disease or history of acute myocardial infarction might create an hypoxic or anoxic condition which, whether temporary or persisting, might result in a subtle but generalized decrement in cortical function.

Lezak (1976) discusses verbal changes that arise with the aging process and refutes the interpretation of previous research that reports immediate memory span (i.e., auditory retention span) diminishes with age. Her own interpretation might well serve to explain how small but statistically significant changes in Token Test performance might result in a population of individuals who begin to show cognitive, rather than verbal, changes which arise with the aging process. She writes:

> Contrary to conventional belief, normal aging processes do not affect the immediate memory span. . . lowered Wechsler Digit Span subtest scores at older age levels result mainly in a greatly shortened span of recall for Digits Backwards, reflecting impaired concentration and mental tracking (page 170).

It seems possible that impaired concentration and mental tracking might account for a somewhat higher error rate among the older subjects who took the Token Test. This is a particularly tantalizing explanation when the redundance of vocabulary and the lack of intrinsic interest value in successive Token Test items is considered.

While all subjects comprising the oldest age group would not display difficulty with concentration and mental tracking, it seems reasonable to suspect that these skills might suffer some reduction for individuals with a history of heart disease, acute myocardial infarction, arteriosclerosis, diabetes, or any other condition which reduced cerebrovascular circulation to some extent. As a higher number of people in the oldest age group are at risk for such problems (Table 7), affected individuals might be expected to generate more errors on a measure like the Token Test when their performance is compared to a group of people who are somewhat freer of such physical conditions. Thus, we might account for the difference in age group performance on the Token Test in this fashion.

Auditory Function and Age-Related Changes on the Token Test

Changes in hearing acuity and discrimination may be one of the major factors which gradually impose limits upon auditory receptive skills of older adults. Hallowell Davis writes that presbycusis is a result of middle ear conductive impairment, as well as the more generally acknowledged sensory-neural changes occurring in the inner ear (Davis & Silverman, 1970). He writes:

> . . .the older a person becomes, the more exposure to noise, with its effects on hearing, does he accumulate. It now seems, however, that even without severe noise exposure or recurrent otitis media elderly people, particularly those beyond 80 years of age, do develop a middle-ear conductive hearing loss. The impairment is greater for high than for low frequencies. Also it is greater for air conduction than for bone conduction. The presence of an "air-bone gap" in the audiogram clearly establishes this part of the hearing impairment to be due to changes in the middle ear.
>
> The nature of the changes in the middle ear with age have not been clearly established, but it is well known that in elderly people connective tissue loses much of its elasticity. Their skin becomes flabby and wrinkled. If such changes occur in the ligaments of the joints between the ossicles so that the bones are not held snugly together, or if the drum membrane loses its stiffness and becomes flabby, there should be just such conductive impairments (page 110).

Davis also writes that a conductive cochlear hearing loss might result with aging, as ". . .loss of elasticity and increase of internal friction in the basilar membrane, that is, a more 'leathery' character, would give the characteristic pattern of gradual high-tone loss. Other tissues show just such changes as part of the aging process" (Davis & Silverman, 1970, page 112).

These sources of hearing loss in the elderly are in addition to those more commonly considered problems of cumulative noise exposure, degeneration of sensory cells in the Organ of Corti, and loss of neurons in the central nervous system (Davis & Silverman, 1970).

How might a reduction in auditory acuity or discrimination result in the increase in error frequency noted among the oldest age group? The subjects, after all, reflected a group of normal, non-diseased adults functioning quite well with reference to daily activities and communicative interactions.

To begin, with, reduced auditory acuity and auditory discrimination might interfere with accurate detection and recognition of connectives found in the final section of the Token Test. Stimuli in this section show the greatest syntactic variability, although the vocabulary critical for accurate comprehension does not possess prominence in the individual TT item's structure. Prepositions, conditionals, conjunctions, adverbs--these word classes, not found in preceding sections of the Token Test are all embedded in the context of verbal commands where they occur as unstressed words, usually a single morpheme in length.

Consider these three items taken from the sixth part of the Token Test:

(i) Touch the blue circle with the red square.

(ii) Touch the blue circle and the red square.

(iii) Pick up the blue circle or the red square.

These three items are administered in consecutive order. Unless the examiner takes special care to emphasize (through increased temporal duration, vocal intensity, and interpretive phrasing) the differences between (i) and (ii), subjects might readily interpret (ii) as a repetition of the preceding item and therefore fail to vary their response to that task. Example (iii) might be modified by the acknowledged change in the verb (and note that it is appropriate to stress the verb in the administration of this item), yet the conjunction 'or' might go overlooked as a contrast to (i) and (ii).

In observing the subjects of all age groups while administering the TT, this process seemed to recur a number of times.

The analysis of error type on the Token Test (Tables 13 and 14) reveals that errors in the Preposition and Other categories account for more than 50% of the errors made by each age group. However, as shown in Table 15, the oldest age group is the one group which ranks these two word categories as the two highest contributing sources of error.

It may be significant that the sixth part of the Token Test follows five sections of relatively unvarying syntactic organization. According to Sanders (1977), as expectancy is developed for stimulus recognition, the amount of attention necessary to process auditory information properly may decline. Sanders states:

> . . .The evidence suggests that by virtue of the structure which the listener imposes upon the incoming acoustic signal, he need perform only a cursory examination of that signal. We perceive according to the probabilities we have used to generate expectancies. These probabilities are computed on the basis of sampling the acoustic information, yet that very sampling is itself influenced by expectancies derived from earlier structure and past experience (pages 156 - 157; emphasized statement is Sanders').

Thus, subjects making the transition to Part F on the Token Test might not be attentive to some of the syntactic changes in initial items because of expectations developed during their experience with former sections of the test. In addition, mild deficits in auditory acuity or discrimination might interfere with their ability to alert themselves quickly (and therefore presumably increase their level of attention) to the more difficult syntactic demands of this final section.

Authors are inclined to minimize phonologic difficulty for Token Test items (DeRenzi and Vignolo, 1962; DeRenzi & Faglioni, 1979). However, the assumption of phonologic ease may hold only for those listeners whose hearing skills remain unaffected by age and environment.

Token Test Performance, Hearing, and Information Theory

Earlier in this discussion an effort was made to attribute age-related reductions in Token Test performance to changes in auditory retention ability. This association had to be tempered, however, as the correlations between Token Test scores and Sentence Repetition scores by each age group were found to be non-significant. Taken as a composite group, however, the .440 correlation between TT and SR performance was found to be significant at the .01 confidence level. Consequently, there appears to be some legitimate leeway to reconsider the question of auditory retention ability and its relationship to Token Test Performance.

Acoustic information is by nature a transitory, non-persisting stimulus. A method of storing this information (i.e., short term memory) is necessary to permit processing of the acoustic pattern in order to derive meaningful comprehension of the stimulus. Short term memory, however, has a finite capacity for information it can store (Sanders, 1977). While auditory memory is limited in terms of its information capacity, the amount of acoustic material which constitutes a unit of information may vary greatly; it is this fact which may offer a perspective for relating age-correlated changes in TT performance with changes in auditory retention ability.

Information is that aspect of a stimulus which allows us to increase the probability of a correct prediction through limitations of perceptual/cognitive choice (Sanders, 1977). Thus, for highly redundant material, a unit of information (or 'chunk') might conform to a phrase rather than to a morpheme. In minimally redundant contexts, an information chunk might conform to a morpheme or even a phoneme. The point is that for normal individuals short term memory capacity has a fixed limit which is measured in terms of information chunks; but the amount of material contained in each chunk will depend upon the nature of the material being presented, the context in which it is presented, and the listener's ability to develop expectancies for forthcoming information based upon his previous experience as a listener.

All subjects participating in this study were normal, healthv adults. Thus, no argument is being advanced that their short term memory systems (i.e., their auditory retention abilities) were deficient in the amount of information retained or processed. However, the possibility exists that for the oldest age group the size of each information chunk may be more restricted as a result of auditory acuity or discrimination changes. If changes in hearing require more careful attention to smaller, unstressed words in the final section of the Token Test, phonologic considerations may shrink the size of each information chunk which is presented for processing. With reduced phonologic certainty, temporal dimensions of phrasing or visual/articulatory cues (as for lip reading) may become more significant than previously.

The Token Test is generally regarded as a test of minimal redundance (DeRenzi & Vignolo, 1962; Boller & Vignolo, 1966; Waller & Darley, 1978; DeRenzi and Faglioni, 1979). This is true if test items are considered in a context that compares commands in the TT to verbal discourse in a more familiar, real life situation. It is also true if redundancy is assessed as a feature which may be present or absent within single stimulus items (i.e., either a given Token Test command is or is not redundant). However, the Token Test consists of a series of 39 sentence-length commands which are given consecutively, rather than on discontinuous occasions. The Token Test is highly redundant if redundance is judged from an overall perspective or from a subtest-to-subtest perspective. A subject who recognizes the internal syntactic consistency for Parts A and B will develop an expectancy that governs listening strategy for Part C.

The problem is that Part F violates many of the expectancies developed during the initial subtests of the Token Test. The structural consistency is no longer as apparent, although the listener is not initially alerted to anticipate a change. The number of stimulus items the listener must respond to within the subtest has increased. The essential vocabulary necessary for accurate execution of the test items increases quite notably, as this is the only subtest where verb recognition is incumbent upon the listener, as well as recognition of prepositions, adverbs, plural markers, conjunctions, conditionals, and negatives.

If, in addition, the listener has the added burden of deficient auditory discrimination (which might aid in accurate recognition of unstressed, non-emphasized critical elements) or deficient auditory acuity, he may continue to chunk information according to strategies previously developed in earlier TT portions but do so while erroneously identifying component morphemes of each chunk. After all, the listener has no previously developed expectation, no reason to anticipate the nature and content of the correct stimulus. On the other hand, the listener might adopt a new strategy for chunking information, where each chunk is comprised of a smaller amount of acoustic material. In this instance, however, additional time might be important to permit accurate processing of the smaller information chunks without loss of continuity with portions of the acoustic message still undergoing verbal presentation.

Thus, it seems a credible conclusion that age-related Token Test scores might be accounted for, at least in part, by changes in auditory retention. These changes, however, are considered to reflect variations in the amount of material contained in each information chunk (the size of the chunk, so to speak), rather than any loss of information retention capacity. It is therefore hypothesized that there is a decrement in the amount of acoustic material contained in each chunk of information. This is concluded to reflect adjustment to an impairment in the peripheral sensory system responsible for relaying acoustic information to the cortical areas where chunking and processing occurs. No compromise of cortical function is therefore inferred on the basis of the above argument.

CHAPTER IV

CONCLUSIONS AND IMPLICATIONS FOR FUTURE RESEARCH

Summary

In this study, the Spreen & Benton version of the Token Test (1969b) was administered to 20 normal, non-brain damaged adults in each of three age categories: 25 through 34 years, 50 through 59 years, and 75 through 84 years. Raw data revealed that the oldest subjects tended to obtain lower scores than the younger two age groups; a simple one-way analysis of variance revealed a significant relationship (at the .01 confidence level) between age and Token Test performance. Subsequently, a Tukey-Snidecor test demonstrated that the performance of the oldest age group was significantly different from the youngest age group and from the middle age group. Differences between the two younger age groups were non-significant.

In addition to the Token Test, the Sentence Repetition subtest from Spreen and Benton's Neurosensory Center Comprehensive Examination of Aphasia (1969b) was administered to the 60 subjects participating in this project. The SR subtest was judged to be a measure of auditory retention ability. A Pearson r correlation was calculated to determine the relationship between performance on the Token Test and performance on the Sentence Repetition subtest of the NCCEA. Correlations were non-significant for the three individual age groups, although the .44

correlation for the entire test population (N = 60) proved to be significant at the .01 confidence level. This was interpreted as an indication that auditory retention is not as critical a factor in TT performance for normal, non-brain damaged adults as might have been anticipated.

An analysis based on error type was attempted for the Token Test. This analysis consisted of tallying the errors made by subjects in each age group and sorting errors into categories descriptive of different word classes (color, shape, size, verb, preposition, other).

The middle and oldest age groups incurred more than 60% of their respective TT errors through faulty recognition of key words falling in the preposition and 'other' categories. Fifty-five (55%) percent of the TT errors made by subjects in the youngest age group likewise were distributed solely within these two error classes. However, when error frequencies for each word class were ranked within age groups, errors in the 'other' word category ranked fourth as a source of Token Test error for the young age group and ranked first as the leading source of error for the remaining two age groups (Table 15). Errors in the preposition category ranked first as a contributing source of error for the youngest age group, ranked fourth as a source of error for the middle age group, and ranked as the second most frequent type of TT error within the oldest age group.

Previous research was reported to have shown TT scores were not influenced by age. In an effort to account for the discrepancy between findings of this present study and a previous one by Orgass and Poeck (1966), mention was made of their small sample size, the extremely limited number of subjects in the oldest age category, and their failure to specify exact ages of the five subjects whose scores were reported in their "60 & over" age group. It was suggested that outcomes of the two studies might differ more in interpretation than in substance, as the data reported by Orgass and Poeck for their oldest age group were in conformance with data obtained in the current study for subjects in the middle age group. Given Orgass and Poeck's failure to report mean ages and age ranges for their oldest age group, it was hypothesized that their "60 & over" age group might have possessed a mean age closer to the middle age group in the present study than to the older subjects reported on in the preceding chapter.

Several possible explanations were offered for the age-related differences in Token Test performance elicited from the three age groups. Age-related diminution of auditory retention skills proved to have a non-significant role in Token Test performance as judged by Pearson r coefficients calculated for each age group on the basis of TT and SR raw scores. It was therefore not possible to account for declining Token Test scores in the oldest age group primarily on the basis of declining auditory retention capacity.

It was argued that with aging comes a greater incidence of hearing loss, diabetes, heart disease, arteriosclerosis and other conditions which might result in a subtle but generalized decrement in cortical function (Weg, 1975; Lezak, 1976). This might account in part for the differences in TT performance between the youngest two age groups and the oldest age group. Minimal changes in mental tracking and concentration skills might account for the noted differences in obtained TT scores.

Changes in auditory acuity and discrimination were discussed as another contributing factor towards the higher TT error rate for the oldest age group.

Finally, information theory was employed as a theoretical framework to explain more specifically how changing hearing acuity and auditory discrimination skills might account for the lower Token Test scores produced by the oldest age group.

Conclusions

As an outcome of the obtained data, it is possible to conclude there is a significant difference in the performance of elderly, non-brain damaged adults when their Token Test scores are compared to scores achieved by younger individuals. There seems to be a logical connection between Token Test performance of older individuals and their individual audiologic status, but to substantiate this statistically would require more rigorous audiologic screening of all potential subjects.

While individual Token Test scores for older subjects show more range, greater variance, and a lower group mean, it is important to recognize that these scores are purely descriptive. That is, the scores describe what may typically be expected from a normal reference group of healthy older adults. Thus, lower Token Test scores in the elderly population do not imply that these individuals are experiencing functional impairment of their receptive language capabilities. In fact, the oldest group's mean for raw Token Test scores was less than 4 points below the group mean obtained by the youngest age group (Table 8).

The Token Test continues to be a useful diagnostic tool for assessing high level impairment in auditory comprehension ability, despite present findings that scores are negatively influenced by age. Current findings simply indicate more caution should be adopted in interpreting possible reasons for depressed Token Test scores for various individuals, particularly those at more advanced age levels.

Studies of Token Test performance might help us to learn more about the language problems experienced by pathological populations. Research in the past has examined Token Test performance with normal adults, with brain-damaged adults, with children. Present data suggest that imperfect test scores may not be mere reflections of singular disorders falling into simple diagnostic categories.

For example, aphasia researchers and aphasia therapists tend to attribute errors on receptive language tasks to cortical language processing deficits. However, if documented changes in peripheral auditory skills could be shown to contribute to greater error rate on the Token Test for normal adults, researchers and clinicians would probably need to conclude that audiologic status as well as aphasic (cortical) language disturbances are responsible for inaccuracies in receptive comprehension demonstrated by patients. If research outcomes begin to document this conclusion, it might lead to more careful audiologic assessment of language-impaired individuals. In turn, this might lead to better identification of intervention strategies to optimize auditory reception.

One final conclusion emerges in the form of a question. If older adults tend to obtain somewhat lower Token Test scores than younger adults, does this possibly indicate these older adults are somewhat more at risk for accurate understanding of verbal legal agreements, verbal warranties, time-compressed radio or television ads, high pressure sales approaches, or rapid and technical explanations of written contracts? Additional research can provide insight into the answers to this question. As we continue to learn more about "normal" language function in all of its variations, we may also begin to acquire more insight regarding the implications these "variations of normal" have for individuals at different life stages in our society.

Implications for Future Research

The present study raises a number of questions that additional research might address. More careful investigation into the role of grammar, syntax, and suprasegmental aspects of the Token Test might prove beneficial. Such a study might consider the question of informational redundancy on an individual item, subtest, and full-test basis.

Future attention might be directed to the roles of expectancy and syntactic consistency in Token Test performance. For example, all 20 tokens might be placed on continuous display, and sentence stimuli from the Token Test might be administered in random order to minimize any expectation of a progressive pattern in successive subtests. An alternative might be to administer the last, least consistent section of the Token Test at the very beginning and then to administer the first five subtests in correct sequence after that.

Future research should attempt to screen subjects more carefully for auditory acuity. If subjects falling into the oldest age category in the present study were divided into two groups on the basis of auditory acuity, it might be possible to determine whether there is a significant correlation between hearing status and TT error rate.

Additional attention should also be directed toward the relationship between positive past medical histories of cardiovascular disease, diabetes, arthritis, chronic hypertension, and auditory reception abilities. Careful testing might indicate that the presence of some of these conditions, rather than advanced age per se, is a contributing factor toward somewhat lower Token Test scores.

Finally, research and carefully deliberated thought should be directed towards identification of situations older adults might be expected to encounter from day to day, where subtle changes in receptive language status might put them at a disadvantage. If we begin to recognize specific circumstances in our society where competent and functional adults are at somewhat higher risk of misunderstanding obligations, rights, instructions, or explanations due to the manner of verbal delivery, we can then collectively move towards ameliorating these occurrences.

APPENDIX

Appendix 1. Stimulus items for Part V of the DeRenzi & Vignolo *Token Test* (1962).

1. Put the red circle on the green rectangle.
2. Put the white rectangle behind the yellow circle.
3. Touch the blue circle with the red rectangle.
4. Touch--with the blue circle--the red rectangle.
5. Touch the blue circle and the red rectangle.
6. Pick up the blue circle or the red rectangle.
7. Put the green rectangle away from the yellow rectangle.
8. Put the white circle before the blue rectangle.
9. If there is a black circle, pick up the red rectangle. (N.B. There is no black circle.)
10. Pick up the rectangles, except the yellow one.
11. When I touch the green circle, you take the white rectangle. (N.B. Wait a few seconds before touching the green circle.)
12. Put the green rectangle beside the red circle.
13. Touch the rectangles, slowly, and the circles, quickly.
14. Put the red circle between the yellow rectangle and the green rectangle.
15. Except for the green one, touch the circles.
16. Pick up the red circle--no!--the white rectangle.
17. Instead of the white rectangle, take the yellow circle.
18. Together with the yellow circle, take the blue circle.
19. After picking up the green rectangle, touch the white circle.
20. Put the blue circle under the white rectangle.
21. Before touching the yellow circle, pick up the red rectangle.

(A. DeRenzi and L. Vignolo, "The Token Test: A Sensitive Test to Detect Receptive Disturbances in Aphasia," *Cortex*, 1962. Reprinted in *Aphasia: Selected Readings*, 1971. Martha Taylor Sarno, editor, 324-325).

Appendix 2. Questionnaire employed in interviewing potential subjects.

Name: ____________ Date of Birth: ____________

Education: ____________ Occupation: ____________

Sex: ____________ Date of Examination: ____________

History:

1.	Is English your primary language?	YES	NO
2.	Is English the first language you learned to speak?	YES	NO
3.	Are you free of recent acute illness?	YES	NO
4.	Have you ever had a stroke?	YES	NO
5.	Have you ever been treated for head injury or for a brain tumor?	YES	NO
6.	Are you in good emotional health?	YES	NO
7.	Have you ever been treated for any disease of the nervous system?	YES	NO
8.	Are you diabetic?	YES	NO
9.	Do you have a history of hypertension (high blood pressure)?	YES	NO
10.	Are you free of kidney or renal disease?	YES	NO
11.	Are you free of color blindness?	YES	NO
12.	Are you free of any visual field problem, such as "tunnel vision?"	YES	NO
13.	Do you wear glasses or contact lenses for reading?	YES	NO
14.	Do you have a hearing loss?	YES	NO
15.	Do you wear a hearing aid?	YES	NO
16.	Do you have any form of seizure disorder (epilepsy)?	YES	NO

Test Situation:

1. Subject wears: ________ Glasses
   ________ Contact lenses
   ________ Hearing aid
2. ________ TT--(3 min.)--SR
   ________ SR--(3 min.)--TT
3. Required formal hearing screening: ________ YES
   ________ NO

(results of hearing screening on reverse of this page.)

Appendix 3. Identification by Sentence (Token Test), Neurosensory Center Comprehensive Examination for Aphasia battery (Spreen & Benton, 1969a). Answer sheet specimen.

Subject No.: _________ Mode. _________ Date: _________

Token Test (TT)

A.	Present all tokens. May repeat instructions once.	**
1.	Show me a circle.	
2.	Show me a square.	
3.	Show me a yellow one.	
4.	Show me a red one.	
5.	Show me a blue one.	
6.	Show me a green one.	
7.	Show me a white one.	
	Total A(7)	

B.	Present only large tokens. May repeat instructions once.	**
8.	Show me the yellow square.	
9.	Show me the blue circle.	
10.	Show me the green circle.	
11.	Show me the white square.	
	Total B(8)	

C.	Present all tokens. No repetition.	**
12.	Show me the small white circle.	
13.	Show me the large yellow square.	
14.	Show me the large green square.	
15.	Show me the small blue square.	
	Total C(12)	

Subject No.: _________ Mode. _________ Date: _________

Token Test (TT)

D.	Present large tokens only. No repetition.	**
16.	Take the red circle and the green square.	
17.	Take the yellow square and the blue square.	
18.	Take the white square and the green circle.	
19.	Take the white circle and the red circle.	
	Total D(16)	

E.	Present all tokens. No repetition.	**
20.	Take the large white circle and the small green square.	
21.	Take the small blue circle and the large yellow square.	
22.	Take the large green square and the large red square.	
23.	Take the large white square and the small green circle.	
	Total E(24)	

F.	Present large tokens only. No repetition.	**
24.	Put the red circle on the green square.	
25.	Put the white square behind the yellow circle.	
26.	Touch the blue circle with the red square.	
27.	Touch the blue circle and the red square.	
28.	Pick up the blue circle OR the red square.	
29.	Move the green square away from the yellow square.	
30.	Put the white circle in front of the blue square.	
31.	If there is a black circle, pick up the red square.	
32.	Pick up all squares except the yellow one.	

Subject No.: _________ Mode. _________ Date: _________

Token Test (TT)

33.	Put the green square beside the red circle.	**
34.	Touch the squares slowly and the circles quickly.	
35.	Put the red circle between the yellow square and the green square.	
36.	Touch all circles, except the green one.	
37.	Pick up the red circle--no--the white square.	
38.	Instead of the white square, pick up the yellow circle.	
39.	Together with the yellow circle, pick up the blue circle.	
	Total F(96)	
	Total A-F (163)	
	Percentile	

Appendix 4. Sentence Repetition (SR), Neurosensory Center Comprehensive Examination for Aphasia battery (Spreen & Benton, 1969a). Specimen answer sheet.

SUBJECT NO.: ____________ Mode: ____________ Date: ____________

<u>Sentence Repetition (SR)</u>

1. Look.	
2. Come here.	
3. Help yourself.	
4. Bring the table.	
5. Summer is coming.	
6. The iron was quite hot.	
7. The birds were singing all day.	
8. The paper was under the chair.	
9. The sun was shining throughout the day.	
10. He entered about eight o'clock that night.	
11. The pretty house on the mountain seemed empty.	
12. The lady followed the path down the hill toward home.	
13. The island in the ocean was first noticed by the young boy.	
14. The distance between these two cities is too far to travel by car.	

page 2 SUBJECT NO.: ____________ Mode: ____________ Date: ____________

Sentence Repetition (SR)

15.	A judge here knows the law better than those people who must appear before him.	
16.	There is a new method in making steel which is far better than that used before.	
17.	This nation has a good government which gives us freedoms not known in times past.	
18.	The friendly man told us the directions to the modern building where we could find the club.	
19.	The king knew how to rule his country so that his people would show respect for his government.	
20.	Yesterday he said he would be near the village station before it was time for the train to come.	
21.	His interest in the problem increased each time that he looked at the report which lay on the table.	
22.	Riding his black horse, the general came to the scene of the battle and began shouting at his brave men.	

BIBLIOGRAPHY

BIBLIOGRAPHY

Bailie, Elaine and Lathrop, Julie, "Development of Norms on the Token Test," unpublished paper, Michigan State University, Department of Audiology and Speech Science, East Lansing, Michigan. December, 1971.

Boller, F. and Vignolo, L. "Latent Sensory Aphasia in Hemisphere Damaged Patients: An Experimental Study with the Token Test," Brain, 89(4), 1966, 815-830.

Davis, H. and Silverman, S.R., Hearing and Deafness, 3rd ed.; Holt, Rinehart and Winston, Chicago, 1970.

DeRenzi, E. and Vignolo, L. "The Token Test: A Sensitive Test to Detect Receptive Disturbances in Aphasia," Brain, 85, 1962, 665-678.

Ebel, R.L., Essentials of Educational Measurement, Prentice-Hall, Inc., Englewood Cliffs, 1972.

Eisenson, J. Adult Aphasia, Appleton-Century-Crofts, New York, 1973.

Gaddes, W.H. and Crockett, D.J., "The Spreen-Benton Aphasia Tests, Normative Data as a Measure of Normal Language Development," Brain and Language, 2(3), 1975, 257-280.

Goodglass, H. and Blumstein, S., Psycholinguistics and Aphasia, Johns Hopkins University Press, Baltimore, 1973.

Lezak, M. Neuropsychological Assessment, Oxford University Press, New York, 1976.

Linton, M. and Gallo, P.S. Jr., The Practical Statistician: Simplified Handbook of Statistics, Brooks/Cole, Monterey, California, 1975.

McNeil, M.R., and Prescott, T.E., The Revised Token Test, University Park Press, Baltimore, 1978.

Orgass, B. and Poeck, K. "Clinical Validation of a New Test for Aphasia," Cortex, 2, 1966, 222-243.

Schuell, H., Jenkins, J.J., and Jimeniz-Pabon, E. Aphasia in Adults, Harper and Row: New York, 1964.

Spellacy, F.J., and Spreen, O. "A Short Form of the Token Test," Cortex, 5(4), 1969, 390-397.

Spreen, O. and Benton, A.L., Neurosensory Center Comprehensive Examination for Aphasia, Answer Booklet, Edition A., University of Victoria, British Columbia, 1969a.

Spreen, O. and Benton, A.L., Neurosensory Center Comprehensive Examination for Aphasia, (Edition A), Manual of Instructions, Neuropsychology Laboratory, Department of Psychology, University of Victoria, British Columbia, 1969b.

Swisher, L.P. and Sarno, M.T., "Token Test Scores of Three Matched Patient Groups," Cortex, 5, 1969, 264-273.

Van Dongen, H.R. and Van Harskamp, R. "The Token Test: A Preliminary Evaluation of a Method to Detect Aphasia." Psychiatria, Neurologia, Neurochirurgia, 75, 1972, 129-134.

Waller, M.R., and Darley, F.L., "The Influence of Context on the Auditory Comprehension of Paragraphs by Aphasic Subjects," Journal of Speech and Hearing Research, 21(4), 1978, 732-745.

Weg, Ruth B., "Changing Physiology of Aging: Normal and Pathological," in Aging: Scientific Perspectives and Social Issues, D. Van Nostrand Co., New York, 1975.

Whitaker, H.A. and Noll, D.J., "Some Linguistic Parameters of the Token Test," Neuropsychologia, 10(4), 1972, 395-404.

Williams, F., Reasoning with Statistics, Holt, Rinehart and Winston, Chicago, 1968.