VERBAL COMPREHENSION AND VERBAL COMMUNICATION IN NEGRO AND WHITE CHILDREN

Thesis for the Degree of M. A.
MICHIGAN STATE UNIVERSITY

Arnold S. Carson

1959

VERBAL COMPREHENSION AND VERBAL COMMUNICATION IN NEGRO AND WHITE CHILDREN

BY

ARNOLD S. CARSON

A THESIS

Submitted to the College of Science and Arts
Michigan State University of Agriculture and
Applied Science in partial fulfillment of
the requirements for the degree of

MASTER OF ARTS

Department of Psychology

1959

Dedication

To My Wife - Lilyan

Acknowledgements

The author wishes to express his sincere gratitude to Dr. Albert I. Rabin who, in his capacity as chairman of this thesis committee, afforded the opportunity for learning from both the student's errors and successes.

Further gratitude is also expressed to Dr. Charles Wrigley for his thorough guidance on statistical analysis and to Dr. Gerald King for his helpful observations and suggestions.

A special note of thanks is directed to Mr. Wayne Maes who graciously arranged the administrative details that were necessary for data collection and to the principals of the schools who assisted the author in finding subjects.

Sincere appreciation is expressed to my wife who assumed all of the responsibility for the typing of the thesis.

VERBAL COMPREHENSION AND VERBAL COMMUNICATION IN NEGRO AND WHITE CHILDREN

BY

ARNOLD S. CARSON

AN ABSTRACT

Submitted to the College of Science and Arts
Michigan State University of Agriculture and
Applied Science in partial fulfillment of
the requirements for the degree of

MASTER OF ARTS

Department of Psychology

1959

Abstract

While testing Southern-born Negro children who had recently migrated to the North, the author of the present thesis was impressed by what seemed to be poor performance in their level of vocabulary responding. It was observed that the same children achieved marked improvement after residing in the North for at least one year.

The investigator reviewed the literature pertaining to the intelligence of the American Negro and found that his observations were in accordance with previous research findings. The question which remained unanswered by the research literature is the extent to which the variables of geographical residency, racial background, and cultural opportunities influence the achievement of Northern and Southern Negro children on tasks requiring a precise and abstract definition.

In the present thesis the Wechsler Intelligence Scale for Children (WISC) Vocabulary Subtest and an adaptation of Rabin, King, and Ehrmann's Qualitative Vocabulary Scale, previously used for the measurement of mental impairment, were utilized to determine cultural differences. It was assumed that measuring devices which are highly discriminatory of cultural differences would also be sensitive to the same differences.

Two sets of hypotheses were phrased in reference to the two experimental measuring devices. It was predicted that Northern White children earn higher raw scores on the WISC Vocabulary Subtest than Northern Negro children and Northern Negro children earn higher raw scores than Southern Negro children. It was also predicted that when the children's responses are scored for seven levels of verbal communication, the Northern White children earn more of higher level scores and less of lower level scores than Northern Negro children, and Northern Negro children earn more of higher level scores and less of lower level scores than Southern Negro children.

Thirty Northern White and thirty Northern Negro children were matched for age, grade, and raw scores on the Full-Range Picture Vocabulary Test with thirty Southern Negro children. Each of the subjects were administered the two experimental vocabulary scales. The Qualitative Vocabulary Scale was scored by two independent scorers to establish sufficient scorer reliability.

The three groups' performance on the WISC Vocabulary Subtest were significantly different in the predicted direction. An overall test of performance on the Qualitative Vocabulary Scale established the three groups' independence. Comparisons of the children's performance on

each of the seven classes of the Qualitative Vocabulary Scale were also computed. Thirteen of the twenty-one comparisons were significant in the predicted direction.

It was concluded from the results that racial intelligence is not a homogeneous entity but instead a variable dependent upon geographical residence and the nature of measuring devices utilized. It was further concluded that the Qualitative Vocabulary Scale is an instrument which is very sensitive to cultural and geographical influences.

Approved <u>A. Raben</u>

Date <u>3/9/59</u>

Table of Contents

List of Tables

Graphs

I Introduction

This study arose from observations on clinical diagnostic findings with Southern Negro children enrolled in the Lansing, Michigan School System. It was attempted to ascertain the cultural differences in verbal communication for these children and comparable Negro and White children from the North. Two independent vocabulary scales were administered, and statistical comparisons were made with the respective scores. Both vocabulary scales were scored in a manner which penalized and rewarded responses for degree of precision and level of abstraction in order to bring forth the differences in verbal facilities among the children.

A secondary purpose was to experiment with the limits of applicability of the qualitative method of scoring vocabulary responses. The scoring method had been used previously for determining mental deterioration. In the present thesis it was utilized as a method of determining cultural differences.

A. The Nature of Racial Differences in Intelligence

Two theoretical viewpoints have confronted each other in the compilation of research on racial differences in intelligence. The first viewpoint assumes biologically innate racial superiority and inferiority in mental heritage. Garth aptly states the position of those psychologists working within this framework:

> Following the geneticist, the psychologist holds to the law of heredity. Theoretically at least he holds it to be true in mental heredity, as well as in morals, that men do not gather grapes of thorns nor figs of thistles. And "ye shall know them by their fruits" is a good maxim in psychology, especially in the above connection (15, p. 25).

Opposition to this first position is stated in the political maxim that "all men are created equal." This viewpoint held research evidence of racial superiority in intelligence dependent upon nurture, selection, and results of careless measuring. After examining the literature to 1931, Garth expressed his personal opinion supporting the second theoretical viewpoint:

> ...the author is convinced after an examination of the literature that we have never, with all our searching, found indisputable evidence for belief in mental differences which are essentially racial (15, p. 24).

Due to the contradiction of the preceding theoretical positions and their inability to account for research

findings in the literature, a third viewpoint arose in the form of a compromise. It contended that racial differences do exist in what the psychologist calls intelligence. However, these differences are not dependent upon a racial inferiority-superiority continuum but instead are a product of poor control in measuring device variables such as language-usage and sampling. Klineberg discussed seven of these variables which must be controlled when studying racial differences. They are 1) Motivation, 2) Rapport, 3) Culture, 4) Social and Economic Status, 5) Language, 6) Schooling, and 7) Sampling. He states;

> The factors we have discussed as playing a part in racial comparisons make clear at least some of the difficulties in the way of basing racial psychology upon tests of intelligence. Until and unless these factors are adequately controlled, such comparisons will have little meaning. It should be added that although no one of these factors accounts by itself for the observed racial differences, the combination of all of them, acting in different ways and to a variable extent depending upon specific circumstances, may very well be responsible (19, p. 177).

B. The Language Variable

Klineberg's fifth variable "Language" was of particular importance for the present thesis. Early in the history of mental testing Terman and Merrill (26) recognized the correlation between language and intelligence.

They stated, "Language essentially is the shorthand of the higher thought processes, and the level at which this shorthand functions is one of the most important determinants of the level of the processes themselves (26, p. 5)."

In accepting this important idea, the psychologist interested in cultural differences is confronted with two contradictions. His measuring devices are standardized in a particular culture while his subjects are more often than not of a culture not participating in the measuring device standardization. The existing mental ability tests are highly verbal and utilize a language which is often not familiar to various racial groups. Klineberg's compilation of research of linguistic and performance tests has clearly shown a trend of foreign groups earning higher IQs on the performance tests than on the linguistic tests (26). In a discussion of this trend he stated, "If we compare the studies of various racial groups by these two types of tests, it becomes clear that the linguistic tests place many of them at a disadvantage (26, p. 169)."

Because of the intimate connection between language and cultural heritage, the highly verbal intelligence test elicits the differences among sub-cultural groups for whom the language is not native. Therefore it may be

an index of cultural opportunities and differences in cultural conditions. It was illustrated in the present thesis that the vocabulary test is very sensitive to cultural differences and may therefore be used diagnostically to determine such conditions and opportunities.

C. Recent Research on the Intelligence of the American Negro

Klineberg administered the National Intelligence Test, Scale A, Form I to 1697 twelve year old Negro boys and girls attending Harlem schools in 1931 and 1932. In order to determine if the superior northern environment had any effect in raising the IQ of Southern-born Negro children, he tested children differing in length of residence in the North. For his control group he used Northern-born Negro children. His results illustrated that intelligence scores improved as length of residency in the North increased. Klineberg concluded from his results:

> The conclusion is therefore justified that the superior showing of those subjects who have had a longer period of residency [in the North] is due to this longer residency and not to any regular change in the quality of migrants (21, p. 31).

Klineberg cross-validated the preceding results by admini-

stering the Binet Intelligence Scale to Southern-born Negro children who resided in the North at lengths varying from one year to eight years and to a control group of New York-born Negro children. He found close but not perfect relationship between test scores and length of residency in New York, improvement taking place most markedly in the first five or six years of Northern residency. The average difference between the Northern-born Negro and Southern-born Negro child was seven Binet IQ points (19).

Coppinger and Ammons (6) collected normative data on the Full-Range Picture Vocabulary Test (F-RPVT), a test not calling for verbal production, writing, or reading. They found that Lousiana Negro children earned mean scores two years lower than a comparable group of White children. They retested the same children with the 1937 Revision of the Stanford-Binet Vocabulary Scale and found that the Negro group had mean scores from three to five years lower than the White group on this measure. The reliability of the F-RPVT was .96 as measured by the correlation of Forms A and B. The test validity coefficient of .84 was based upon a correlation with the 1937 Revision of the Stanford-Binet Vocabulary Scale. Coppinger and Ammons concluded that common norms are justified for

White and Negro children when these children share the same culture or when the Negro children have adopted the White children's cultural standards. They suggested that special test norms should be adopted when making intra-Negro group comparisons since the variability is high among Negro children and dependent upon the cultural standards adopted (6). The findings of the Coppinger and Ammons study indicated that non-verbal measures of Negro intelligence were less biased and at the same time less sensitive to cultural and racial differences.

Klineberg appropriately summarized what may be expected when testing the American Negro on Verbal Intelligence:

> The case of the American Negro is somewhat different. His native tongue is English; it is usually the only language he speaks. In many cases, however, he speaks it so badly that he is placed at a great disadvantage in intelligence testing involving anything like a precise discrimination of the meaning of terms (19, p. 168).

It is again emphasized that it is the verbal aspect of the intelligence test which is sensitive to linguistic differences and therefore contains the potentiality for a measuring device of racial differences.

D. The Vocabulary Scale As a Measure of Intelligence

1. Rationale for usage. The vocabulary scale is one of the oldest and most reliable determinants of intelli-

gence. Its validity depends upon the framework within which intelligence is defined. It is worth noting that while intelligence is defined in various ways there are very few actual intelligence tests which do not include a vocabulary subtest. Terman and Merrill stated their opinion about the vocabulary scale which is in the 1937 Revision of the Stanford-Binet,

> We have found the vocabulary test to be the most valuable single test in the scale. Its interest value is high, it presents a familiar task to the subject, and the fact that it begins with words in common use and increases rapidly in difficulty gives the examiner a rapid survey method of estimating the subjects's ability. It agrees to a high degree with the mental age rating on the scale as a whole; correlations for single age groups range from .65 to .91 with an average of .81 (26, p. 302).

Watson summarized the rationale for the Vocabulary Subtest of the Wechsler-Bellvue Intelligence Scale:

> It would appear to be simultaneously a measure of the patient's fund of information, range of ideas, and learning ability....
>
> Rapaport states that performance is to a considerable degree dependent upon the wealth or poverty of early educational influences and is comparatively unmodified by later schooling and life experiences....
>
> Wechsler and Rapaport agree that vocabulary is quite refractory to impairment. This results in part from acceptance in the scoring system with equal credit of different and conceptually lower levels of definitions. Thus it does not necessarily mean that there is no change in verbal functioning following impairment or deterioration. Actually the individual may function

> verbally at a somewhat lower level than would have been the case when his vocabulary was at its peak of scope and efficiency and yet manage to give barely satisfactory definitions.... (28, pp. 180-182).

The present investigator is not in full agreement with Rapaport's contention about the permanent influence of early educational influences upon the vocabulary response. Klineberg's studies (19, 20, 21) on qualitative changes in the vocabulary response of Southern-born Negro children residing in the North during later childhood and adolescence are inconsistent with Rapaport's contention. It should be noted that both the 1937 Revision of the Stanford-Binet and the Wechsler-Bellvue Intelligence Test use a correct or incorrect scoring system which does not measure higher or lower levels of verbal communication. The present thesis capitalized on the fact that a vocabulary response reflects different levels of verbal communication varying in precision of discrimination and level of abstraction. The preceding features of the vocabulary response demanded that it be scored in a manner which penalizes and rewards according to the qualitative difference of the response.

2. <u>History and evolution of the qualitatively scored vocabulary response</u>. In 1904 Chambers experimented with scoring vocabulary responses of young children according

to four gradations of correctness (4). Almost three decades later Green (17) weighted vocabulary responses according to five degrees of precision and discovered that the subjects' total weighted scores correlated higher with their IQs than did their quantitatively scored vocabulary responses.

In 1940 Yacorzynski (32) published his hypothesis in connection with an evaluation of the postulates underlying the Babcock Deterioration Test. Yacorzynski contended that the often noted stability of vocabulary test scores in mental deterioration may be more apparent than real because scoring is based on an "all or none" principle and is independent of qualitative differences in responding. The hypothesis proved to be fruitful since it gave impetus to approximately 15 research studies which in turn proved the validity of the hypothesis.

Table 1 charts the evolution of the qualitative scoring system for vocabulary responses and describes the categories into which the responses were classified. Whenever the information was available, the measuring devices of the studies were also reported. It should be noted that the 1952 Chodorkoff and Mussen study (5) was the only one which failed to validate Yacornzynski's hypothesis in that it found no significant differences between the responses of 40 normal and 40 schizophrenic

subjects. The present investigator attributed the lack of significant results to the insensitivity of their multiple choice measuring device. When subjects are given a choice of responses, they may choose the best sounding response without origination from their own volition. It was therefore further contended that the Chodorkoff and Mussen study measured the capacity of a subject to recognize the most accurate vocabulary response and did not measure the capacity of the subjects originally to choose or to communicate a response.

Table 1 illustrates that there is some disagreement among the current research workers as to whether the "descriptive" or the "functional-usage" type response is of higher value. The investigator designated the descriptive type of response as a higher level of verbal communication than the functional-usage type response on the basis of research evidence from two studies conducted by Stacey and Portnoy (25). Using Gerstein's suggested method of analysis (16) for a comparison of normal adults and children with mentally retarded adults and children, Stacey and Portnoy illustrated that the descriptive response is a higher level of conceptualization than the functional-usage response. A further safeguard was added to the scoring system used in the present thesis by adding the adjectives "essential" and "vague" to the appropriate

Table 1

Evolution of Qualitative Scoring System for Vocabulary Responses

Date	Authors	Measuring Device	Methods of Classification
1904	Chambers (4)		1. no answer 2. wholly wrong answer 3. vaguely right 4. correct
1931	Green (17)	Stanford-Binet Vocabulary	1. synonym 2. use & description 3. explanation 4. illustration, demonstration, repitition 5. error
1940	Yacornzynski (32)	Hypothesis: The often noted stability of vocabulary test scores in mental deterioration may be more apparent than real because scoring is based on "all or none" principle and independent of qualitative differences in the responses.	
1944	Watts (29)		Abstract-Concrete Continuim with 1. labels for concrete - lowest 2. universal & abstract - highest
1949	Feifel (12)	Stanford-Binet	1. synonym 2. use, description, & use and description 3. explanation 4. illustration, inferior explanation, repitition, demonstration 5. error

Table 1 Continued

1949	Gerstein (16)	W-BIS[a] Vocabulary	1. categorical or conceptual 2. functional or usage 3. concretistic or descriptive
1952	Chodorkoff & Mussen (5)	Multiple Choice S-H[b] Retreat Scale	1. class 2. description 3. example 4. function
1954	Harrington & Ehrmann (18)	Standard & Multiple Choice Adm. of W-BIS[a] Vocabulary	4 levels of abstraction
1955	Rabin, King & Ehrmann (23)	FRPVT[c], Paw Score Vocabulary W-BIS[a], Verbal Derivation of FRPVT[c]	1. synonym or definition 2. functional-descriptive 3. concretistic 4. partially irrelevant 5. completely irrelevant
1955	Ehrmann (9)	W-BIS[a] Vocabulary	1. class or categorical 2. descriptive 3. example 4. functional 5. error

a. W-BIS: Wechsler-Bellvue Intelligence Scale

b. S-H: Shipley-Hartford Retreat Scale

c. FRPVT: Full-Range Picture Vocabulary Test

categories. It is contended that if a scoring system is to be truly qualitative, it should account for precision vs. vagueness as well as abstractness vs. concreteness.

E. The Qualitative Scoring System in the Present Thesis

The present qualitative scoring system is a sevenfold arrangement consisting of three superior response classes and four inferior response classes. Although a decision has been made as to which classification represents the highest level of verbal communication, the second highest, and so forth to the lowest level, there is no interval measurement used with this system. Since the intervals between classifications may not be equidistant, all scores are in the form of tallies.

Traditionally the "categorization" and "synonym" type responses were considered the most superior levels of verbal communication. These classifications have been given equal scoring weight and comprise the highest rated class. Because of the research findings of Stacey and Portnoy (25), the "essential descriptive" response was designated as the second highest level of verbal communication and the "essential function" response as the third highest level. The ordering of the four lower levels of verbal communication was based on the consistent

trend in the evolution of the qualitative scoring system as represented in the literature to score the "example" response higher than either the "vague description" or "vague function" response.

The experimenter has attempted simultaneously to account for two variables in the scoring of responses: 1) level of abstraction and 2) degree of precision. A description of the seven classes (from highest to lowest) and sample responses which would be tallied in the respective classifications follows:

Class 1 - CATEGORIZATION AND SYNONYM

a. Categorization: The categorization response classifies by some definite scheme in terms of its universal characteristics.

Stimulus Word - "wagon" Response - "a vehicle"

b. Synonym: The synonym response may essentially be used to replace the object or idea with no or little change in the denotative aspects of the stimulus word.

Stimulus Word - "wagon" Response - "a cart"

Class 2 - ESSENTIAL DESCRIPTION: An essential description response must give the characterizing features of the stimulus word. If the stimulus word is abstract, the response must create mental imagery of the emotional situation. If the stimulus word is concrete (physically

tangible), the response must differentiate between the stimulus word and members of its class.

Stimulus Word - "wagon" Response - "It's a wooden thing with four wheels, and it looks like a box."

Class 3 - ESSENTIAL FUNCTION: An essential function response must describe primary rather than peripheral usage or purpose of an object or an idea.

Stimulus Word - "wagon" Response - "It's pulled by a horse and you ride in it out west."

Class 4 - EXAMPLE: An example response defines an object or idea in terms of its aspects or members.

Stimulus Word - "wagon" Response - "There's a red wagon kids play with."

Class 5 - VAGUE DESCRIPTION AND VAGUE FUNCTION

a. Vague Description: A vague description is a response that is not totally irrelevant but does not give the characterizing features of the object or the idea.

Stimulus Word - "wagon" Response - "Something that has four wheels."

b. Vague Function: A vague function response describes the peripheral rather than primary usage or purpose of the object or the idea.

Stimulus Word - "wagon" Response - "It bumps into people."

Class 6 - ERROR: The error response is totally irrelevant to the stimulus word.

Stimulus Word - "wagon" Response - "The dog wagons his tail"

Class 7 - DON'T KNOW: A "don't know" response is a statement or a lack of statement given by the subject designating that he is unable to verbally define a word which he had previously designated recognition for by a non-verbal response.

Stimulus Word - "wagon" Response - " don't know"

The present investigator found it unfeasible to separate the "Vague Description" and "Vague Function" type responses into two independent classifications which would correspond to the separation of the "Essential Description" and "Essential Function" type responses. The caliber of the "Vague" responses was so poor that it was often difficult to differentiate the functional from the descriptive characteristics of the response. A few sample stimulus words and responses will help illustrate this difficulty:

Stimulus Word	Response
1. steel	"It's real tough to bend"
2. firecracker	"bang paper"
3. hot	"It's warm"

F. Purpose and Hypotheses

The purpose of the present study is to test if Northern White children manifest higher levels of verbal communication than Northern Negro children and to test if Northern Negro children manifest higher levels of verbal communication than Southern Negro children.

The hypotheses in the present thesis pertained specifically to the measuring devices. Hypotheses 1 through 3 were phrased with reference to the classification used in conjunction with the qualitative scoring system. Hypotheses 4 through 6 were phrased with reference to a standardized vocabulary scale which was scored according to the standardized instructions.

When the three groups are matched for non-verbal word recognition:

1. Northern White children earn greater frequencies of classes 1, 2, and 3 type responses and lesser frequencies of classes 4, 5, 6, and 7 type responses than Southern Negro children.

2. Northern Negro children earn greater frequencies of classes 1, 2, and 3 type responses and lesser

frequencies of classes 4, 5, 6, and 7 type responses than Southern Negro children.

3. Northern White children earn greater frequencies of classes 1, 2, 3 type responses and lesser frequencies of classes 4, 5, 6, and 7 type responses than Northern Negro children.

4. Northern White children earn higher raw scores on a standardized vocabulary scale than Southern Negro children.

5. Northern Negro children earn higher raw scores on a standardized vocabulary scale than Southern Negro children

6. Northern White children earn higher raw scores on a standardized vocabulary scale than Northern Negro children.

II Description of Materials

The F-RPVT (2) was constructed and validated as a vocabulary test not calling for verbal response, writing, or reading but instead a test of gesticulation as a means of communicating word recognition. It was utilized in the present study as a measuring device for matching the three groups of children on their ability to identify words without verbal communication. The test's normative sample consisted of 480 White children, 120 White adults, 80 Spanish American children, and 80 Negro children. The authors of the test claim alternate-form reliability ranging from the low .80's to the high 90's for the various age levels.

The test materials consist of 16 plates, each plate containing four cartoon-like drawings, and a list of 170 words, each word associated with a particular drawing. In order to respond during testing, a subject points or gestures to the drawing on the plate which best illustrates a given word. The explicit value of the test is suppose to be the measurement of word recognition uninfluenced by verbal expression.

One of the two experimental measuring devices was the Wechsler Intelligence Scale for Children (WISC) Vocabulary Subtest. The vocabulary subtest yields split-

half reliability coefficients ranging from .77 to .90 and correlation coefficients with the full scale WISC ranging from .63 to .83 for three age groups: 7½, 10½, and 13½ years old (31).

The second experimental measuring device was a revision of a qualitative vocabulary scale by Rabin, King, and Ehrmann (23). The subjects were requested to define verbally the same words which they had previously been tested on for matching purposes (F-RPVT). The children were given the opportunity to define only those words which they correctly identified through gesticulation behavior in previous testing. The purpose of determining a cut-off point in the above manner was to allow the matched pairs of children equal opportunities in the number of verbal word definitions. All responses were scored according to the proposed qualitative scoring system discussed earlier.

III Subjects and Procedure

A. Subjects and Method of Matching

The subjects were Southern Negro (SN), Northern Negro (NN), and Northern White (NW) children enrolled in the fourth, fifth, and sixth grades. None of the subjects were advanced or behind in their age-grade placement. The entire group of 90 subjects were drawn from four schools in the Lansing School System. The SN group comprised all of the Negro children within the age range from 9.5 to 11.5 years who were born and reared in the South and have migrated to Lansing, Michigan within 28 months of the date tested. All of the NW and NN children were born and reared in Ingham County, Michigan. Children were tested with the F-RPVT at random from the latter two groups in order to acquire subjects whose scores could be matched with the scores of the children in the SN group. Table 2 summarizes in detail the relevant matching data for the three groups of children. The groups, consisting of 30 subjects each, were matched on grade, age, and raw scores on the F-RPVT.

Table 2

Grade, Sex, Age, and Matching Data of the Three Groups

		Grade			Sex		Ages (Months)		Raw Score of F-RPVT	
Group	N	4th	5th	6th	Male	Female	Mean	SD	Mean	SD
NW	30	16	8	6	16	14	128	10.9	43.67	4.97
NN	30	16	8	6	15	15	125	10.9	43.57	5.74
SN	30	16	8	6	14	16	128	13.5	43.53	5.21

B. Administration of the Tasks

The F-RPVT and the two experimental measuring devices were administered individually to the subjects by the same examiner in single testing sessions. The SN children were tested first. Subjects were then chosen randomly from the NW and NN populations and tested with the F-RPVT. Random testing among the two populations was discontinued when each SN child was matched with a NW child and a NN child on the basis of their scores on the F-RPVT. The order in which the tests were administered and the average time of administration follows.

First - F-RPVT: 7 minutes

Second - WISC Vocabulary Subtest: 13 minutes

Third - Qualitative Vocabulary Scale: 18 minutes

C. Scorer Reliability

Two independent scorers tallied the responses of eight protocols from each of the sample groups in accordance with an instruction sheet which they read initially and had access to throughout scoring. (See Appendix A). The twenty-four protocols were chosen from the total sample of 90 protocols with the help of a table of random numbers (27). The task of the scorers was to tally each response into one of six classifications on the basis of the definition of the class and sample responses for each class as they were presented on the instruction sheet. There was no scorer reliability check for class 7 since the placing of tallies in this class did not involve any scorer judgement.

In order to obtain a measure of scorer reliability the number of agreements between the two scorers was totaled for each of the respective classes. The number of tallies designated in each class by the scorers were added together in order to obtain the total number of tallies. The formula $\frac{2(\text{number of agreements})}{\text{total number of tallies}}$ was used to derive the percentage of agreement between the two scorers. The results of the preceding computations follows:

Class 1 - percentage of scorer agreement = 89%

Class 2 - percentage of scorer agreement = 82%

Class 3 - percentage of scorer agreement = 81%

Class 4 - percentage of scorer agreement = 79%

Class 5 - percentage of scorer agreement = 72%

Class 6 - percentage of scorer agreement = 83%

In order to obtain the total percentage of agreement between scorers all of the agreements were totaled and the total number of tallies given by each of the scorers was added together. The preceding formula was again utilized to measure the total percentage of scorer agreement; in this case it was 79%.

Complete agreement between the two scorers was obtained after the debatable responses had been discussed. Discussion of the debatable responses suggested that the disagreement between the two scorers was a result of not thoroughly exploring the qualitative features of all the responses. To be more explicit, there was a tendency to classify responses by their similarity to previously scored responses rather than the criteria of the instruction sheet.

IV Results and Discussion

A. Results

Every subject was classified as either a higher or lower level of communication responder. The higher level respondents were those subjects who earned a greater number of tallies in classes 1, 2, and 3 while the lower level responders were those subjects who earned a greater number of tallies in classes 4, 5, 6, and 7.

The chi-square statistic was utilized to test whether the number of low level scorers and high level scorers in each sample group departed significantly from chance when compared with the other two sample groups. The computed chi-square in the present thesis was 22.54, significant beyond the .01 level.

Table 3 presents the relevant mechanics of the chi-square computations. Column I lists the seven sample-level contingencies. Column II lists the actual number of subjects per sample classified as a higher or lower level responder. Column III lists the corresponding theoretical expectancies per cell, and Column IV illustrates the contribution of each sample-level contingency to the total chi-square.

Table 3

The Chi-Square on the Higher and Lower Levels of Verbal Communication

Level and Sample	Frequencies		Contribution to Total X^2
Column I	Observed Column II	Expected Column III	Column IV
High - NW	25	16.7	4.13
High - NN	18	16.7	.10
High - SN	7	16.7	5.63
Low - NW	5	13.3	5.69
Low - NN	12	13.3	.12
Low - SN	23	13.3	7.07
			Total X^2 22.54
			df = 2
			$p < .01$

All of the hypotheses were tested with the Rank Test for Paired Observations (24). As the title of the statistic might suggest, the scores in each of the groups are ranked from lowest to highest and the difference between the ranks of the paired observations are computed.

Hypothesis 1: NW children earn greater frequencies of classes 1, 2, and 3 type responses and lesser frequencies of classes 4,5,6, and 7 type responses than SN children.

Table 4 presents a comparison of the two groups' responses which were classified according to the qualitative scoring system described earlier. The predicted differences were upheld with the exception of the differences in classes 4 and 5. NW children did not earn lesser frequencies of "Example" or "Vague Description and Vague Function" type responses than SN children.

Table 4

Comparison of the NW and SN Children

On Classes 1,2,3 and 4,5,6,7

	Ranks		Theoretical			Level of
Class	NW	SN	Mean	SD	z	Significance
1	271.0	29.0	150.0	35.00	3.44	$p < .0005$
2	327.5	107.5	217.5	46.25	2.36	$p < .01$
3	350.0	85.0	217.5	46.25	2.85	$p < .0005$
4	274.5	160.5	217.5	46.25	-1.22	n.s.
5	154.5	223.5	189.0	41.62	.82	n.s.
6	34.5	316.5	175.5	39.37	3.57	$p < .0005$
7	3.0	432.0	217.5	46.25	4.63	$p < .0001$

Hypothesis 2: NN children earn greater frequencies of classes 1, 2, and 3 type responses and lesser frequencies of classes 4,5,6, and 7 type responses than SN children.

Table 5 presents a comparison of the two groups' responses which were classified into the respective categories. The results of comparisons on classes 1, 3, 6, and 7 substantiated the prediction. However, NN children did not earn greater frequencies of the "Essential Description" type response, class 2, or lesser frequencies of the "Vague" type response, class 5, than SN children. NN children earned significantly greater instead of significantly lesser "Example" type responses, class 4, than SN children which was contrary to the hypothesis.

Table 5

Comparison of the NN and SN Children

On Classes 1,2,3 and 4,5,6,7

	Ranks		Theoretical			Level of
Class	NN	SN	Mean	SD	z	Significance
1	181.0	72.0	126.5	30.80	1.76	$p < .05$
2	206.5	171.5	189.0	41.62	.41	n.s.
3	326.0	139.0	232.5	48.62	1.91	$p < .05$
4	301.0	105.0	203.0	43.91	2.22	$p < .025$*
5	202.5	203.5	203.0	43.91	.00	n.s.
6	55.5	295.5	175.5	39.37	3.04	$p < .0025$
7	40.5	394.5	217.5	46.25	3.82	$p < .0005$

* The difference was significant in the direction contrary to the hypothesis.

Hypothesis 3: NW children earn greater frequencies of classes 1,2, and 3 type responses and lesser frequencies of classes 4,5,6, and 7 type responses than NN children.

Table 6 presents a comparison of the two groups' responses which were classified into the respective categories. Four of the seven predicted differences were supported. NW children did not earn significantly greater frequencies of "Essential Function" type responses or significantly lesser frequencies of "Example" and "Vague Description and Vague Function" type responses than the NN children.

Table 6

Comparison of the NW and NN Children

On Classes 1,2,3 and 4,5,6,7

	Ranks		Theoretical			Level of
Class	NW	NN	Mean	SD	z	Significance
1	289.0	62.0	175.5	39.37	2.87	$p < .0025$
2	283.5	122.5	203.0	43.91	1.82	$p < .05$
3	233.0	118.0	175.5	39.37	1.45	n.s.
4	126.5	198.5	162.5	37.15	.82	n.s.
5	147.0	231.0	189.0	41.62	1.00	n.s.
6	57.5	218.5	138.0	32.88	2.43	$p < .01$
7	95.5	229.5	162.5	37.15	1.79	$p < .05$

Table 7 summarizes the results of comparisons obtained from the children's performance on the Qualitative Vocabulary Scale. Fourteen of the 21 predictions relating to this scale were supported by the results. Seven of the nine predictions relating to superior levels were upheld, and seven of the twelve predictions were supported for inferior levels. In examining Table 7 it may be noted that the comparisons on the highest (class 1) and the lowest (classes 6 and 7) classifications manifest the greatest certainty as to the significant differences between the groupings.

Table 7

Summary of Levels of Significane Obtained from Comparisons on the Qualitative Vocabulary Scale

	Classes of Response						
	Superior			Inferior			
Comparisons	1.	2.	3.	4.	5.	6.	7.
NW vs. SN	.0005	.01	.005	n.s.	n.s.	.0005	.0001
NW vs. NN	.0025	.05	n.s.	n.s.	n.s.	.001	.05
NN vs. SN	.05	n.s.	.05	.025*	n.s.	.0025	.0005

* The difference was significant in the direction contrary to the hypothesis.

Hypothesis 4: NW children earn higher raw scores on a standardized vocabulary scale than SN children.

Hypothesis 5: NN children earn higher raw scores on a standardized vocabulary scale than SN children.

Hypothesis 6: NW children earn higher raw scores on a standardized vocabulary scale than NN children.

Table 8 presents the comparisons that were made in order to test Hypotheses 4, 5, and 6. It is evident from the significance of the results that each of the hypotheses was supported.

Table 8

Comparisons on the WISC Vocabulary Subtest

Comparisons	Ranks			Theoretical			Significance
	NW	NN	SN	Mean	SD	z	
NW vs. SN	432.0		3.0	217.5	46.25	4.62	$p < .0001$
NN vs. SN		345.5	60.6	203.0	43.92	3.24	$p < .001$
NW vs. NN	334.0	101.0		217.5	46.25	2.51	$p < .01$

In order to gain a better understanding of the degree of association between simple word recognition and the verbal communication of word recognition, two correlational analyses were computed. The degree of association between the WISC Vocabulary Subtest and the F-RPVT was correlated for each of the three groups. Table 9 presents the results of the correlations.

Table 9

Pearsonian r Between the Raw Scores and Between the MAs of the WISC Vocabulary and the F-RPVT

	Correlations	
Group	Raw Scores of WISC Vocabulary and F-RPVT	MAs of WISC Vocabulary and F-RPVT
NW	.736	.750
NN	.806	.653
SN	.784	.741

B. Discussion

The results of the thesis demonstrated that NW children of comparable non-verbal word recognition abilities manifest higher levels of verbal communication than NN children, and, in turn, the NN children manifest higher levels of verbal communication than SN children.

The reversal of the significant differences between NN and SN children in the comparison of their "Example" type responses pointed out the difficulty in making predictions concerning the middle range between the highest and lowest levels of verbal conceptualization. The predictions concerning the highest and lowest levels of verbal conceptualization were supported with a great deal more certainty than the predictions concerning the middle range of verbal conceptualization.

Graph 1 illustrates the pattern of responding for each of the groups when their vocabulary responses are scored according to a qualitative system. The NW children favor the higher levels of verbal communication. Their percentage of responses decreases sharply in the "Example" classification, and after a slight increase in the "Vague" classification, their percentage of responses continues to decrease lower than the two Negro groups in the "Error" and "Don't Know" classifications. NN children favor the "Vague" type response and manifest a greater percentage

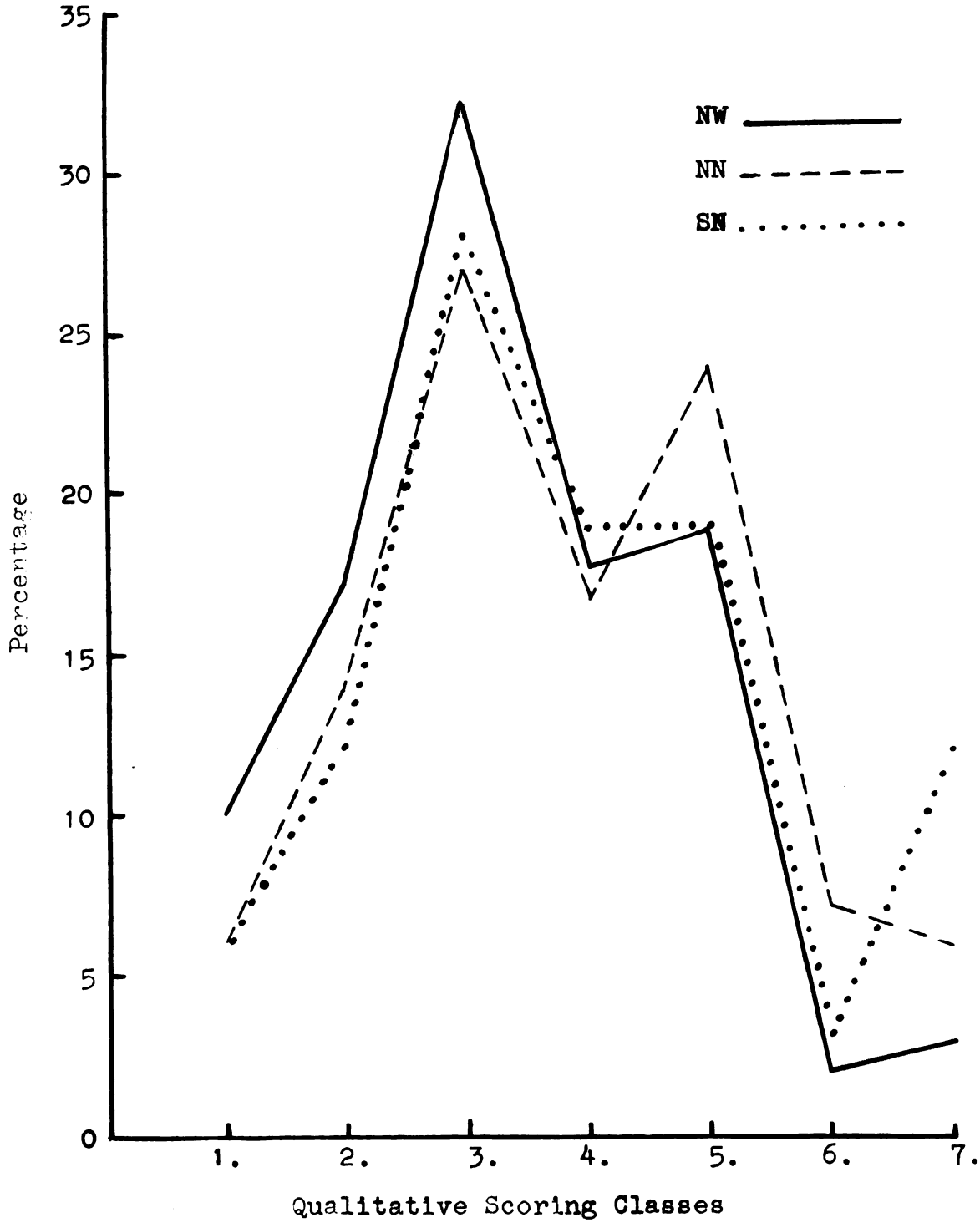
of "Error" type response than the White children. The SN children's pattern is the only one which manifests a high peak in the "Don't Know" classification.

Yacorzynski's (32) hypothesis concerning the complexity of the vocabulary response is incidentally validated in the present thesis by the differences obtained with a seven-class multiple scoring system.

The differences between the groups cannot consistently be explained according to one theoretical viewpoint. The differences between NW and NN children could lend support to either the contention that constitutional racial differences in intelligence exist or the viewpoint that a difference of cultural opportunities accounts for the variance between two racial groups. The NW and NN samples in the present thesis are offsprings of fathers with similiar occupational levels which usually is indicative of little socioeconomic differences. Table 10 presents a D.O.T. (34) (Part IV) Classification of the fathers' occupations. The differences among the fathers' occupational levels for the two groups are not significant. However, there are sufficient possibilities of differences in the cultural milieu of the two samples to question their equivalence of social opportunities. The two grops remain segregated from each other in activities other than those

Graph 1

Pattern Description of Vocabulary Responses
Given by NW, NN, & SN Children
And Scored by a Qualitative 7-Class System



which revolve around school. The NN sample, for the most part, lives in a section of town which is a homogeneous Negro settlement. Their cultural milieu is definitely not characteristic of middle class modes of living while the NW children do live according to middle-class standards or are striving to do so.

Table 10

D.O.T. Part IV Classifications of Occupations

For the NW and NN Children's Fathers*

	D.O.T. (Part IV) Classifications	NW Fathers (N)	NN Fathers (N)
0-x	Professional, Technical, and Managerial Work	2	0
1-x	Clerical and Sales Work	4	5
2-x	Service Work	3	2
4-x	Mechanical Work	9	5
6-x	Manual Work	12	18

* If the father was deceased, the main wage earner was classified instead.

The comparisons of NN and SN children lend themselves to a more clear-cut interpretation since the cultural variable of geographical residence was explicitely controlled. It is contended that the superior educational

environment and the greater opportunity for cultural advancement of the NN child over the SN child accounts for the significant differences between the two groups.

The present study supports Klineberg's (21) research findings that geographic residency of the Negro child is an important determinant of vocabulary performance since there were significant differences in the vocabulary responses of SN and NN children. Coppinger and Ammons' (6) contention that different norms should be utilized when making inter-Negro group comparisons where the members of the groups have different cultural backgrounds is also supported by the significant differences between the SN and NN groups.

V Summary and Conclusions

In order to explore the effect of cultural and measuring device variables on racial intelligence, NW, NN, and SN children were administered two vocabulary tests which called for an expression of verbal conceptualization in responding. It was expected that these highly linguistic measures of intelligence would elicit definite patterns of responding for each of the three samples even though such a test is known to be highly biased as a measure of intelligence.

Six predictions were made which related specifically to both the WISC Vocabulary Subtest and the Qualitative Vocabulary Scale. Generally it was expected that NW children would define more words with greater precision and higher levels of abstraction than NN children, and, in turn, NN children would define more words with greater precision and higher levels of abstraction than SN children.

The results substantiated 14 of the 21 predictions for the comparisons of performance on the Qualitative Vocabulary Scale and all the comparisons of performance on the WISC Vocabulary Subtest. It was concluded that:

1. Differences in racial intelligence can easily be brought about by manipulation of the measuring device variables.

2. Racial intelligence may not be a homogeneous entity, but instead dependent upon variables such as geographical residency, language, and cultural opportunities.

3. The vocabulary test which penalizes and rewards for levels of verbal conceptualization manifests the qualities of a diagnostic tool for determining racial differences.

Bibliography

1. Altus, Grace T. Relationship between verbal and non-verbal parts of the CTMM and WISC. J. consult. Psychol., 1955, 19, 143-144.

2. Ammons, R. B., & Ammons, H. S. The full-range picture vocabulary test. Amer. Psychologist, 1949, 4, 267-268.

3. Atchison, O. Use of the WISC with 80 mentally defective Negro children. Amer. J. ment. Defic., 1955, 60, 378-379.

4. Chambers, W. G. How words get meaning. Ped. Sem., 1904, 11, 30-50. Cited in Ehrmann.

5. Chodorkoff, B., & Mussen, P. Qualitative aspects of the vocabulary responses of normals and schizophrenics. J. consult. Psychol., 1952, 16, 43-48.

6. Coppinger, N. W., & Ammons, R. B. The Full-Range Picture Vocabulary Test: VIII. A normative study of Negro children. J. clin. Psychol., 1952, 8, 136-140.

7. Corter, H. M. Factor analysis of some reasoning. Psychol. Monogr., 1952, No. 8.

8. Dixon, J., & Massey, F. J. Introduction to Statistical Analysis. (2nd ed.) New York: McGraw-Hill, 1957.

9. Ehrmann, J. C. A qualitative analysis of vocabulary responses in early and late maturity. Unpublished Ph.D. thesis, Michigan State University, 1955.

10. Estes, Betsey W. Influence of socioeconomic status on WISC: an exploratory study. J. consult. Psychol., 1953, 17, 58-62.

11. Estes, Betsey W. Influence of socioeconomic status on WISC: addendum. J. consult. Psychol., 1955, 19, 225-226.

12. Feifel, H. Qualitative differences in the vocabulary response of normals and abnormals. Psychol. Monogr., May 1949, 39, 151-204.

13. Feifel, H., & Lorge, I. Qualitative differences in the vocabulary responses of normal children. J. educ. Psychol., 1950, 41, 1-18.

14. Fry, T. C. The x^2 test of significance. J. Amer. statist. Ass., 1938, 33, 513-525.

15. Garth, T. R. Race Psychology A Study of Racial Mental Differences. New York: Whittlesey House, McGraw-Hill, 1931.

16. Gerstein, Reva A. A suggested method for analyzing and extending the use of Bellevue-Wechsler vocabulary responses. J. consult. Psychol., 1949, 13, 366-370.

17. Green H. J. A qualitative method for scoring the vocabulary test of the new revision of the Stanford-Binet. Unpublished M. A. thesis, Stanford University, 1931. Cited in Ehrmann.

18. Harrington, R., & Ehrmann, J. C. Complexity of response as a factor in the vocabulary performance of schizophrenics. J. abnorm. soc. Psychol., 1954, 49, 362-364.

19. Klineberg, O. Race Differences. New York: Harper, 1935.

20. Klineberg, O. (Ed.) Characteristics of the American Negro. New York: Harper, 1944.

21. Klineberg, O. Negro intelligence and urban residence. In T. M. Newcomb & E. L. Hartley (Eds.), Readings in Social Psychology. New York: Henry Holt, 1947. Pp. 24-32.

22. Moran, L. J., Moran, F. A., & Blake, R. R. An investigation of the vocabulary performance of schizophrenics: II. Conceptual level of definitions. J. genet. Psychol., 1952, 80, 107-132.

23. Rabin, A. I., King, G. F., & Ehrmann, J. C. Vocabulary performance of short-term and long-term schizophrenics. J. abnorm. soc. Psychol., 1955, 50, 255-258.

24. Siegel, S. Nonparametric Statistics for the Behavioral Sciences. New York: McGraw-Hill, 1956.

25. Stacey, C. L., & Portnoy, B. A study of the differential responses on the vocabulary subtest of the Wechsler Intelligence Scale for Children. J. clin. Psychol., 1950, 6, 401-403.

26. Terman, L. M., & Merrill, M. Measuring Intelligence. Boston: Houghton Mifflin, 1937.

27. Wallis, W. A., & Roberts, H. V. Statistics A New Approach. Glencoe, Illinois: The Free Press, 1956.

28. Watson, R. I. The Clinical Method in Psychology. New York: Harper, 1951.

29. Watts, A. F. The Language and Mental Development of Children. London: Harrap, 1944.

30. Webster's New Collegiate Dictionary. (1st ed.) Springfield, Mass.: G. & C. Merriam, 1951.

31. Wechsler, D. Wechsler Intelligence Scale for Children. New York: The Psychological Corporation, 1949.

32. Yacorzynski, G. K. An evaluation of the postulates underlying the Babcock deterioration test. Psychol. Bull., 1940, 37, 425-426.

33. Young, Florence M., & Bright, H. A. Results of testing 81 Negro rural juveniles with the WISC. J. soc. Psychol., 1954, 39, 219-226.

34. Division of Occupational Analysis. Dictionary of Occupational Titles Part IV. Washington, D.C.: U.S. Government Printing Office, 1944.

Appendix

Scoring Instructions for the Qualitative Vocabulary Scale

This is a seven-fold scoring system for vocabulary responses. Class 1 responses are the highest scores possible, and Class 7 responses are the lowest scores possible. Choose the class in which a response should be tallied in strict accordance with the Definition of Classes listed below. When a response qualifies without doubt for placement in more than one class, tally it in the higher or highest of these classes. A list of sample responses are provided for each of the classifications.

Definition of Classes

1. CATEGORIZATION AND SYNONYM

a. Categorization: The categorization response classifies by some definite scheme in terms of its universal characteristics.

Sample Responses

1) steel - a metal
2) horse - an animal
3) pie - a dessert
4) telephone - way of communication

b. Synonym: The synonym response may essentially be used to replace the object or idea with no or little change in the denotative aspects of the stimulus word.

Sample Responses

1) dwelling - a place to live
2) clean - sanitary
3) currency - money
4) insect - bug
5) shrubbery - bushes
6) surf - waves of water

2. ESSENTIAL DESCRIPTION: An essential description response must give the characterizing features of the stimulus word. If the stimulus word is abstract, the response must create mental imagery of the emotional situation. If the stimulus word is concrete (physically tangible), the response must differentiate between the stimulus word and members of its class.

Sample Responses

1) counter - when you buy stuff, you put it on it
2) skill - like if somebody is real good in something

3) farm - place where they have animals and grow crops
4) firecracker - a small amount of explosives
5) accident - when you get hit by a car

3. ESSENTIAL FUNCTION: An essential function response must describe primary rather than peripheral usage or purpose of an object or an idea.

Sample Responses

1) pie - something to eat
2) furniture - you can sit on it
3) horse - you can ride him
4) clothes - something you wear

4. EXAMPLE: An example response defines an object or idea in terms of its aspects or members.

Sample Responses

1) athletes - players
2) liquid - coffee, Coka Cola
3) nutrition - food, vitamins
4) recreation - basketball
5) container - cup

5. VAGUE DESCRIPTION AND VAGUE FUNCTION

a. Vague Description: A vague description is a response that is not totally irrelevant but does not give the characterizing features of the object or the idea.

Sample Responses

1) plant - something you plant outside the house
2) farm - place where you have animals
3) steel - it's hard
4) hot - you can't touch it

b. Vague Function: A vague function response describes the peripheral rather than primary usage or purpose of the object or the idea.

Sample Responses

1) farm - it makes things
2) furniture - to play games on
3) plant - it prettys up
4) telephone - used to dial

6. ERROR: The error response is totally irrelevant in defining the stimulus word.

Sample Responses

1) wagon - the dog wagons his tail
2) furniture - the door bell
3) steel - plastic
4) hot - cold

7. DON'T KNOW: A "don't know" response is a statement or a lack of statement given by the subject designating that he is unable to verbally define a word.

Sample Responses

1) currency - don't know
2) telephone - can't say

Commonly Used Abbreviations

L = like
sbdy = somebody
sg = something
DK = don't know

ROOM USE ONLY