1974

An Investigation of videotape rescoring of certain subtests of the Porch Index of Communicative Ability

Caroline Lee Massarotto
The University of Montana

AN INVESTIGATION OF VIDEOTAPE RESCORING OF CERTAIN SUBTESTS OF THE PORCH INDEX OF COMMUNICATIVE ABILITY

By

Caroline Lee Massarotto

Diploma in Speech Therapy, University of Queensland, Australia, 1968

Presented in partial fulfillment of the requirements for the degree of

Master of Arts

UNIVERSITY OF MONTANA

1974

Approved:

Chairman, Board of Examiners

Dean, Graduate School

June 7, 1974
Date

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

ACKNOWLEDGEMENTS

The writer would like to thank her thesis committee members, Dr. Robert Chaney and Dr. Wesley Shellen for their suggestions and advice.

Gratitude is also expressed to the clinicians and patients from the Speech Pathology Departments at the Glenrose Hospital in Edmonton and the Veteran's Administration Hospital in Denver, who went out of their way to help in this study.

Finally, the writer wishes to express sincere appreciation both to Paul and to Dr. Kellogg Lyndes for their help and encouragement. It was a combination of the former's patience and the latter's continued advice that made the completion of this thesis possible.

TABLE OF CONTENTS

Chapter		Page
I.	INTRODUCTION	1
	Statement of the Problem	3
	Review of the Literature	5
II.	PROCEDURE	11
	Subjects	11
	Equipment	12
	Standard Test Conditions	12
	Experimental Procedures	12
III.	RESULTS	15
	Average Scores	15
	Score Changes	18
	Types of Changes	23
	Subjects and Clinicians	26
IV.	DISCUSSION	27
	Analysis of Average Scores	27
	Score Changes	28
	Type of Response	30
	Clinicians	32
	Conclusions	33
V.	SUMMARY	36
BIBLIOGRAPHY		38
APPENDIX		41

LIST OF TABLES

Table		Page
1.	The PICA Categories for Scoring Responses	2
2.	Differences Between Average Scores for VGS Under the Administration and Review Conditions	16
3.	Differences Between Average Scores for Eight Gestural Subtests Under the Administration and Review Conditions for Each Test Given	16
4.	Differences Between Average Scores for Four Verbal Subtests Under the Administration and Review Conditions	17
5.	Comparison of Average Score Changes (Disregarding Sign) for Each Subtest Under the Administration and Review Conditions	19
6.	Standard Deviation of Changes Between the Administration and Review Conditions for Each Subtest for All Tests Given	20
7.	A Comparison of Direction and Mean Size of Score Changes from Administration to Review of Verbal and Gestural Subtests for Twelve Subjects	22
8.	Distribution of Numbers of Changes over 120 Items from the Response Level Given During Administration to that Given During Review for Twelve Subjects	25

LIST OF FIGURES

Figure — Page

1. Total Number of Changes Made for Each of the Twelve Subjects Tested, Related to His/Her Average Score for VGS 21

2. Percentage of Response Changes from that Assigned During Administration at Each of the Levels Used, Over All Items and All Tests Given 24

CHAPTER I

INTRODUCTION

The Porch Index of Communicative Abilities (PICA) is a "clinical tool designed to assess and quantify certain verbal, gestural, and graphic abilities" (Porch, 1973, p. 1). It was further stated by Porch (1973) that the test enables the clinician to measure levels of output ability and make inferences about the input and integrative abilities of the aphasic patient.

The PICA consists of 18 subtests: 4 verbal, 8 gestural, and 6 graphic. It employs a multidimensional scoring system that quantifies 16 levels of responses in terms of varying degrees of accuracy, responsiveness, completeness, promptness, and efficiency. A single score is assigned to a given response which describes the response in varying degrees of these five dimensions. This multidimensional system is viewed as a binary choice system in which the scorer, in evaluating a response from a patient, makes a series of binary choices continuing through a system until a decision yields a score (Porch, 1971). (See Appendix A.) The PICA categories for scoring responses are presented in Table 1.

Porch (1973, p. 1) stated that a "critical factor in the usefulness of the PICA is the reliability of the tester." He went on to say that once the tester has achieved reliability in his administration

TABLE 1

THE PICA CATEGORIES FOR SCORING RESPONSES

Score	Category	Dimensional Characteristics
16	Complex	Accurate, Responsive, Complex, Prompt, Efficient
15	Complete	Accurate, Responsive, Complete, Prompt, Efficient
14	Distorted	Accurate, Responsive, Complete or Complex, Prompt, Distorted
13	Complete-Delayed	Accurate, Responsive, Complete or Complex, Delayed
12	Incomplete	Accurate, Responsive, Incomplete, Prompt
11	Incomplete-Delayed	Accurate, Responsive, Incomplete, Delayed
10	Corrected	Accurate, Self-corrected
9	Repeated	Accurate, after instructions are repeated
8	Cued	Accurate, after cue is given
7	Related	Inaccurate, almost accurate
6	Error	Inaccurate attempt at the task item
5	Intelligible	Comprehensible but not an attempt at the task item
4	Unintelligible	Incomprehensible but differentiated
3	Minimal	Incomprehensible and undifferentiated
2	Attention	No response, but patient attends to the tester
1	No Response	No response, no awareness of task

Source: Porch Index of Communicative Ability--Administration, Scoring, and Interpretation, Vol. 2, p. 17, by Bruce E. Porch (1973).

and scoring, the subtests are sensitive to small changes in the patient's performance.

Reliability can be defined as the relative absence of errors of measurement in a measuring instrument (Kerlinger, 1973). This definition therefore includes terms such as the stability, dependability, predictability, and accuracy of the measure. The concern of this study centered around the stability and dependability of scores assigned by an administrator of the PICA to responses produced by a patient during administration of the test. The method of examination employed the review through the use of videotape recording of the original evaluation during which the administrator rescored the test.

Statement of the Problem

Because of the numerous and frequently rapid judgments required of the test administrator during the conduct of this test, considerable concern was experienced by the writer regarding the ongoing accuracy of the administrator of the test in scoring a continuous stream of responses by the patient. Subsequent rescoring by not only the author but by other trained administrators, through the media of videotape recordings, appeared to support this concern. It became a routine procedure within the University of Montana, Speech, Language, and Hearing Clinic, as well as at the Speech Pathology Clinic at the Glenrose Hospital, Edmonton, Alberta, to review assigned scores by means of video and/or audio tape. Hence, a study of this area was felt to be warranted to explore the usefulness of such a procedure.

These same test administrators in both clinics had reached a level of proficiency that resulted in high inter-observer reliability at the completion of their training on the PICA. Due to such possible factors as infrequent use of the test, speed at which many responses need to be recorded, and the fleeting nature of many responses which must be scored, this proficiency level was subsequently questioned. Either the audio or video review was felt to supplement the proficiency level. The videotape was felt to be most helpful in that it allowed for both auditory and visual cues to be reviewed in the rescoring process.

From empirical observation, it appeared that the review of the videotape supplemented initial scoring for the following reasons: (1) it allowed for repeated observations during the videotape review of a particular response, (2) the fact that the videotape picture was usually taken from a front view of the patient, thus allowing for a view of the patient from a different vantage point than that obtained by the clinician during the original administration, and (3) the elimination of the administrator's concern for standard presentation of instructions, which then allowed for more attention to be given to the patient's response.

The videotape review was found to be most helpful for gestural and verbal subtests because responses were often fleeting. It appeared that graphic subtest scores could be reviewed without videotape for all factors except delay, self-correction, repetition of instructions, and cued instructions. In administration of the graphic subtests the clinician can make note of this reduced number of factors and thus it

was felt that these subtests can be rescored without the necessity of video replay.

The central question asked in this study was whether there was a substantial difference in scores obtained on the verbal and gestural subtests during administration of the test and scores obtained when the test was reviewed through video recordings by the original administrator.

Review of the Literature

Porch (1967) presented as part of the construction of the test an investigation of scorer agreements. The method involved a comparison of the PICA scores obtained by two trained observers and the administrator of the test. Thirty patients were used for this part of the study and a comparison of subtest means and response levels were made. Results indicated that differences between the means of scores were very small (.25 of one point for subtests and .19 of one point for graphic, gestural, verbal and overall response levels) when the potential range of variation (15.0) was considered. This writer suggests that although the possible range of variation is 15 points, the greatest probable range would be 10 points; e.g., the change from a 4 type response (unintelligible) to a 14 type response (distorted). However, even if the difference between means of scores is compared with this probable range, it might still be considered small.

An analysis of subtest score differences showed the following:

1. That no verbal subtest showed significant differences between scorers.

2. The significant variances were found mainly in Subtest III, a gestural subtest, and Subtest E, a graphic subtest.
3. The lack of scorer agreement was attributed to a bias on the part of one of the scorers and this was not consistent over all subtests.

Porch (1971) stated that a given scorer might be inclined to score slightly, though consistently, high or low on various subtests, but that the main threat to scorer agreement was an occasional failure to apply the rules of scoring.

Porch (1971, p. 791) also stated that "the patient's behavior dictates which scoring category must be used to describe that behavior, and therefore, the scorer has little opportunity to allow subjectivity to enter into his decisions." In addition, he said that although a few of the 180 items recorded during the administration of the test allow some freedom of interpretation and subjectivity, when and if differences occur, they were small enough to have little effect on the subtest mean. He concluded by saying:

> Of course, any differences between scorers, no matter how small, mean one clinician is observing differences in behavior that the other clinician is not observing and therefore clinicians who use the system feel a great responsibility to exert every effort to score with precision and sensitivity. It is a system in which clinical standards are far more stringent than psychometric standards (Porch, 1971, p. 791).

The PICA battery of subtests was designed to sample a variety of communicative skills at different levels of complexity. The range of task difficulty was found to be wide, varying from Subtest A (graphic), which was the most difficult task, to Subtest XI (gestural), one of the easiest and least sensitive.

In the test construction, 150 patients representing varying degrees of communicative involvement were used. The scores within the subtests were found to be widely dispersed for this group, covering almost the entire range of the 16-point scale. Broad dispersion of scores was also evidenced in the overall scores and across all modalities (verbal, graphic, and gestural) (Porch, 1967).

The use of videotaping and films versus live presentation in past literature has been compared by this author, in an attempt to determine the efficiency of one type of presentation over the other. Results are inconsistent and vary according to the goals of the study.

The comparisons that have been made cover a wide variety of subjects. These range from studies in which the subjects were tested on the content learned under the two conditions (Taylor, Lipscomb, and Rosemier, 1969), the effectiveness of videotape recordings in attitude change versus a live presentation of the same materials (Wall and Boyd, 1971), to the value of using a videotape monitor in a standard psychological interviewing situation (Denoff, Stenmark, and Smith, 1970).

Taylor, Lipscomb, and Rosemier (1969) studied the difference among three groups. One group interacted with the teacher for three one-hour sessions per week for eight weeks, another group watched these sessions via video tape and the third viewed the first two video tapes of the first group's interaction and then interacted with the teacher during one session for each of the eight weeks. Results indicated that there was no difference in content learned between the live interaction and the videotaped interaction groups. Denoff, Stenmark, and Smith (1970) looked at the difference between videotape presentation

of standard comments in a psychological interview and a face-to-face interview using the same standard comments. Temporal data were compared, e.g., frequency and duration of talking, by the subjects. They found that normal subjects responded equally in terms of this temporal data to the television monitor as to the live interviewer. The Wall and Boyd (1971) study will be discussed further in the next paragraph.

Although there are studies which show contradictory results, the emphasis of these studies has been considerably different from that of this study. Croft, Stimpson, Ross, Bray, and Bieglic (1969) felt that videotape presentations have fewer cues available than the live presentation. The study described these lost cues as including fineness of facial features, color, three dimensionality, quality of voice, and total situational gestalt. Croft et al. (1969) studied attitude change and found that more change occurred with live presentation of a message. It was noted that credibility of the source was felt to be an important factor in effecting attitude change. This study was replicated by Wall and Boyd (1971) who found no difference between the two conditions and suggest that random assignment of subjects to cells was not done by Croft et al. (1969), which may have biased the results.

Flynn (1970) examined the development and evaluation of videotaped discrimination training programs. The subjects were trained to make discrimination judgments of selected phonemes in a training program for the /s/ phoneme which was presented on videotape. The subjects were given pre- and post-tests of their discrimination ability for /s/ʃ/r/, with part presented via videotape and part presented through

live stimulus materials. Similar scores were obtained from both these methods of testing, thus suggesting that the two conditions of videotape and live presentation are equal even for fine discrimination of phonemes.

Cullinan and Prather (1968) compared the effectiveness of rating stuttering by live communication and by audiotape. They then compared their results with those obtained by Williams, Wark, and Menifie (1963), who compared ratings of stuttering by audio and audiovisual clues and suggested that equally reliable rating of the severity of stuttering can be obtained from live rating sessions and audiovisual recordings. The Cullinan and Prather (1968) study, in rating the stutterer, included the following items: the degree of tension and/or associated movements specific to dysfluency; relative duration of dysfluency; relative frequency of dysfluencies; overall severity of stuttering; normality of "nonstuttered" speech; bodily movements not specific to dysfluency; and communication interaction. Each of the items was rated on a scale of increasing abnormality from one to seven.

Videotaping has been used extensively with the intention of detailed analysis of the data for clinical use. Boone and Prescott (1971) presented a speech and hearing therapy scoring manual, the purpose of which was to quantify the interaction between two people or between a clinician and a small group in therapy. The behavior of the clinician and client are categorized in terms of content and sequence; e.g., the clinician describing and explaining the specific goals or procedures of the session would be classed as a category I response. The client making an incorrect response was categorized as 7. The

authors suggest that this system may be used to categorize therapy for any kind of speech therapy problems including articulation, hearing, language, and voice. Videotape replay of a therapy session was recommended for this purpose.

Videotape recordings have also been used to assess articulation progress and as an aid in articulation therapy (Burkland, 1967), as well as a means of recording behavior in free play settings (Norquist, 1971).

In view of the above studies, and the fact that the responses scored on the PICA usually involve obvious movements and verbalizations by the patient, it seems unlikely that the use of the videotape recordings to present the review will cause, in itself, a significant difference in scores between the live and review scoring conditions. That is, the live scoring condition and the videotape scoring condition may be thought of as equal, and any differences found in scores under the two conditions will be due to factors other than live versus videotape presentation.

Reliability studies on standardized speech and language tests as well as psychological tests include inter-tester reliability or test-retest reliability studies, but not intra-tester reliability on the same test. It was felt by the writer that the similarities of test and review procedures outlined above would tend to support, at least in part, the intra-tester procedures employed in this study.

CHAPTER II

PROCEDURE

Subjects

Subjects were speech pathologists who were employed at the Glenrose Hospital in Edmonton, Alberta, Canada, and at the Veteran's Administration Hospital in Denver, Colorado. The subjects were currently using the PICA as part of their diagnostic-therapeutic program. All subjects had fulfilled the standard number of hours of training in the administration of the test. These are stated by Porch (1973) as forty hours of training which should be used to familiarize the scorer thoroughly with the scoring method and manual, as well as practice administration of the battery and scoring responses until the scoring categories and test format are fully memorized. Following this, the testing of at least ten patients with varying degrees of communicative impairment was carried out by each of the clinicians.

Six clinicians, three from each of the above centers, were used. Each clinician tested two patients. Clinician experience varied from a clinician who had taught numerous PICA workshops to one who had learned the test only six months previously. All clinicians learned the test at a PICA workshop conducted either by Bruce Porch or by one of his students and, on the average, had been using the test for at least 18 months.

Equipment

Videotaping equipment, including the camera, recording devices (Sony #3650) and monitor, was borrowed from the setting in which that part of the study was carried out.

The camera was set up to obtain the fullest and clearest picture of the patient being tested without violating the standard test conditions, but allowed for a complete picture of the patient's body and movements, the clinician, and the test items. The microphone was set as close as possible to the patient to allow for adequate reception, without violating the standard test conditions. The microphone placement allowed for clear reception of all the patient's vocal responses, and the clinician's instructions.

Standard Test Conditions

The conditions outlined in PICA, Volume 2, Administration, Scoring and Interpretation, pages 2-11, were followed.

Experimental Procedures

The clinicians did not know the nature of the study nor did they know that they would be asked to rescore the original test.

1. The following verbal instructions, prior to testing, were given to each clinician. "If you are willing to take part in the study, I would like you to administer the PICA to two people and score as you would normally. Your sessions will be videotaped. I would like you to leave two 1½ hour slots open in your schedule on either _____ or _____. The details of the study will be available and I will answer any questions you may have about it after the test

procedures have been completed. Please sit to the right and at right angles to the patient."

2. The score sheets were collected from the clinicians following the testing. They did not calculate any of the subtest means.

3. The review occurred after 48 hours had elapsed but before 72 hours had elapsed following the initial administration. This procedure was employed because it seemed that in the usual clinic setting, the interval between the administration and review would approximate this time interval. This range of time was also necessary because of scheduling difficulties encountered in the various clinical settings. It was expected that there would be some recall of the 120 instances of scoring for each of the tests. The following verbal instructions were given prior to the review: "This is the videotape of your session with M_____, on _____. I want you to re-score the test as you see it now. I will stop the tape any time you need extra time, and you can review the tape any time you are unsure of what response occurred."

4. At the completion of both administrations and reviews, this writer interviewed each of the clinicians. They were given copies of their test and retest score sheets. The clinicians were asked to examine these sheets and if they changed some of their scores, they were asked whether they recalled the reasons for the change. In those instances where the clinician recalled a reason for the change, these reasons were recorded. The clinicians were also asked not to discuss the study nor the content of this interview until all the procedures involved in the study were completed in his or her clinic.

5. Explanation of the purpose of the study was given to the subjects and all questions they had were answered at the completion of the study. When the results of the study have been calculated, copies of the results will be forwarded to each clinician.

6. Each clinician was asked to fill in a questionnaire following both administrations. The questionnaire covered information concerning the experience and training of the clinicians in the administration and scoring of the PICA, whether the patient had been tested previously, and the emotional state of the patient during the administration of the test. (See Appendix B.)

CHAPTER III

RESULTS

The data for this study consisted of PICA test results from the gestural and verbal subtests (VGS), for 12 patients, obtained during administration and, again, during review via videotape. Six clinicians were used, each testing two patients.

Average Scores

The potential range of variation for the following scores was 15 points, but the probable range, as noted earlier, was 10 points. The average score for all of the 12 subjects (VGS) given to the patients employed in this study is presented in Table 2. It should be noted that the differences for each of the tests under the two conditions ranged from 0 to .18 of one point.

The mean scores for the eight gestural subtests for the 12 patients under the two conditions are presented in Table 3. The differences range from .01 to .36, a slightly larger range than for the average of 12 subtest scores reported previously.

Results of the mean scores for the four verbal subtests for the 12 patients under the two conditions are presented in Table 4. The range of differences for these subtests was .03 to .50, which was slightly larger than the range obtained for the gestural subtests as well as for the VGS.

TABLE 2

DIFFERENCES BETWEEN AVERAGE SCORES FOR VGS UNDER THE ADMINISTRATION AND REVIEW CONDITIONS

Clinician	Patient	Administration	Review	Difference
1	a	14.18	14.15	-.03
	b	13.58	13.60	.02
2	c	13.80	13.82	.02
	d	13.38	13.41	.03
3	e	13.42	13.39	-.03
	f	12.28	12.28	.00
4	g	10.34	10.29	-.05
	h	11.64	11.46	-.18
5	i	4.28	4.38	.10
	j	8.29	8.28	-.01
6	k	13.03	12.95	-.08
	l	11.62	11.48	-.14

TABLE 3

DIFFERENCES BETWEEN AVERAGE SCORES FOR EIGHT GESTURAL SUBTESTS UNDER THE ADMINISTRATION AND REVIEW CONDITIONS FOR EACH TEST GIVEN

Clinician	Test	Administration	Review	Difference
1	a	14.25	14.24	-.01
	b	13.83	13.84	.01
2	c	13.91	13.85	-.06
	d	13.54	13.65	.11
3	e	13.24	13.35	.11
	f	13.46	13.43	-.03
4	g	13.96	13.88	-.12
	h	13.86	13.50	-.36
5	i	5.08	5.19	.11
	j	9.23	9.09	-.14
6	k	13.45	13.31	-.14
	l	11.95	12.00	.05

TABLE 4

DIFFERENCES BETWEEN AVERAGE SCORES FOR FOUR VERBAL SUBTESTS UNDER THE ADMINISTRATION AND REVIEW CONDITIONS

Clinician	Test	Administration	Review	Difference
1	a	14.03	13.98	-.05
	b	13.10	13.15	.05
2	c	13.58	13.75	.17
	d	13.05	12.93	-.12
3	e	13.78	13.48	-.30
	f	9.90	10.00	.10
4	g	3.10	3.13	.03
	h	7.20	7.25	.05
5	i	2.67	2.75	.08
	j	6.43	6.65	.22
6	k	12.18	12.23	.05
	l	10.95	10.45	-.50

It is suggested that the greater range of differences found under the two conditions, VGS and verbal subtests, might have been contributed to by the greater number of subtests involved in obtaining the average VGS scores. That is, with the obtained VGS averages, 12 subtest scores contributed to the mean score obtained. In the case of the verbal subtests mean score, only four tests contributed. This might suggest that the greater number of subtest scores may have lent themselves to a more normal distribution. In the case of the gestural subtests, this same argument might also be extended. It should be noted that the greatest range of differences was found between the verbal subtest means (.03 to .50 of a point) with the least range being

found between the VGS means (0 to .18 of a point). The gestural means fell between these two scores (.01 to .36 of a point).

Score Changes

The score changes for each of the subtests were calculated from the difference between subtest means under the administration and review. (See Appendix C.) These were totaled for the 12 subjects without regard to sign. That is, the sum of the differences between subtest means for each of the subtests was calculated. The total number of score changes for each subtest for all 12 tests given was found. (See Appendix D.) The average score change for each subtest for all tests given was obtained by dividing the total number of score changes for each subtest into the sum of differences for each subtest. These results are presented in Table 5. From this table, the following features appeared salient: (a) the average score change for each subtest varied from .25 to .08 of one point; (b) the total number of changes for each of the subtests for all subjects ranged from 39 to 2; (c) the size and total number of changes for the verbal subtests tended to be greater than for the gestural subtests:

Average Sum of Differences (disregarding sign) for Four Verbal Subtests . . . . . = 4.03

Average Sum of Differences (disregarding sign) for Eight Gestural Subtests . . . . . = 2.38

Average Number of Changes for the Four Verbal Subtests . . . . . = 26.75

Average Number of Changes for the Eight Gestural Subtests . . . . . = 16.87

TABLE 5

COMPARISON OF AVERAGE SCORE CHANGES (DISREGARDING SIGN) FOR EACH SUBTEST UNDER THE ADMINISTRATION AND REVIEW CONDITIONS

Subtest	Sum of Differences	Total Number of Changes	Average Score Change
I	5.5*	38*	.14
II	5.0	37	.14
III	5.6	39	.14
IV	4.8*	29*	.17
V	4.2	24	.18
VI	.8	8	.10
VII	1.5	18	.08
VIII	.6	3	.20
IX	3.9*	27*	.14
X	.8	4	.20
XI	.5	2	.25
XII	1.9*	13*	.15

* Verbal subtests

In general, the verbal subtests also showed greater dispersion of size of score changes than the gestural subtests (see Table 6). These data were calculated from Appendix C.

Average Standard Deviation for Verbal Subtests = .48

Average Standard Deviation for Gestural Subtests = .33

Figure 1, showing the total number of changes made for each patient tested, related to his/her average score for VGS, appeared to

TABLE 6

STANDARD DEVIATION OF CHANGES BETWEEN THE ADMINISTRATION AND REVIEW CONDITIONS FOR EACH SUBTEST FOR ALL TESTS GIVEN

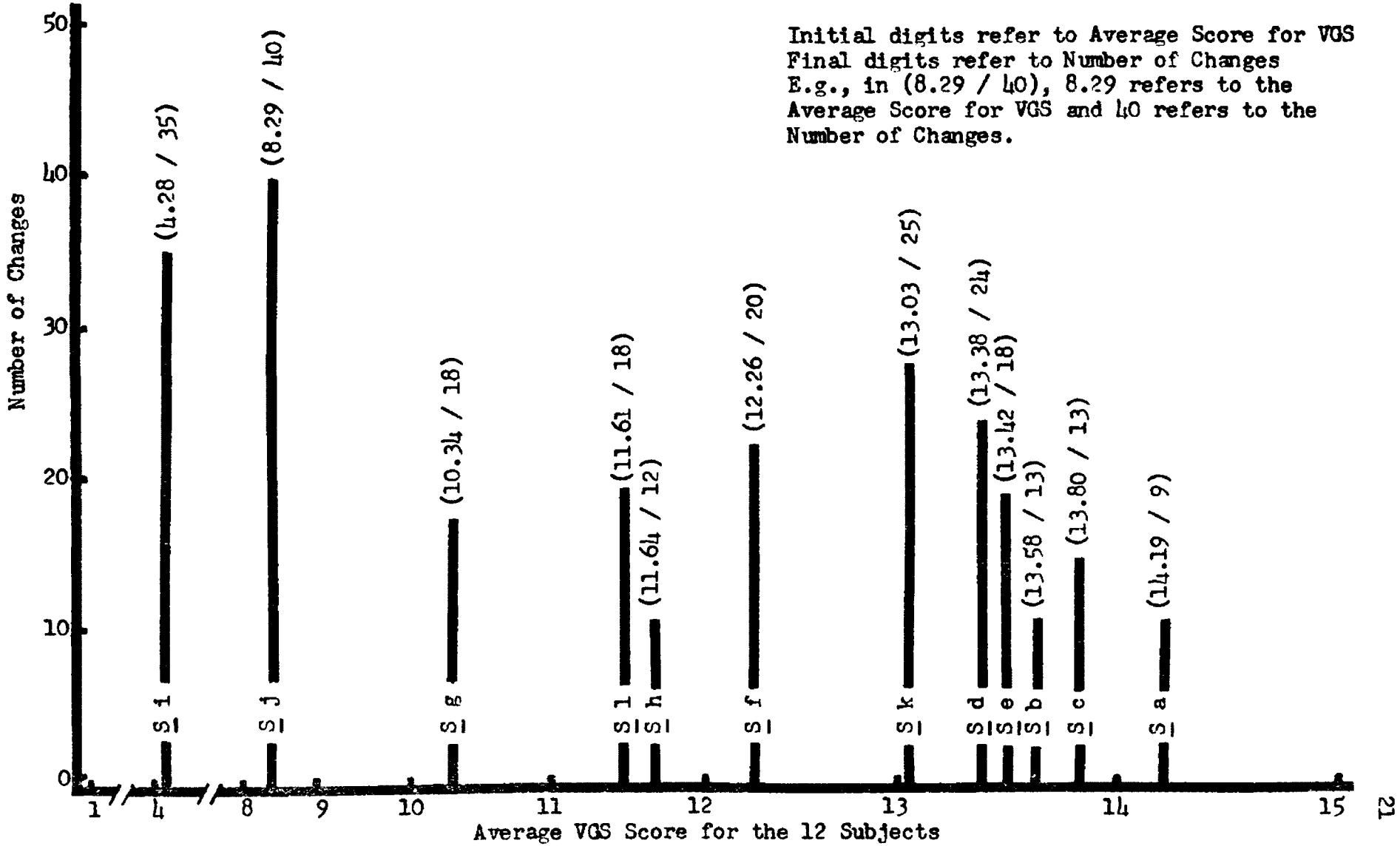
Subtest	Standard Deviation
I	.61*
II	.37
III	.61
IV	.64*
V	.45
VI	.17
VII	.26
VIII	.16
IX	.41*
X	.49
XI	.10
XII	.26*

* Verbal subtests

suggest that the lower the level of overall functioning, the greater the number of score changes that occur. However, a number of cautionary factors associated with this finding need to be noted: (1) Only two of the 12 patients tested functioned below an overall level of 10 for VGS; (2) the same clinician tested both patients who functioned below this level; (3) the only patient whose emotional state was judged by his clinician to have significantly interfered with the administration and scoring of the test was the one whose average for the 12

FIGURE 1

TOTAL NUMBER OF CHANGES MADE FOR EACH OF THE TWELVE SUBJECTS TESTED, RELATED TO HIS/HER AVERAGE SCORE FOR VGS



subtests was 8.29. Forty scoring changes, the highest number throughout the study were made between the two test conditions for this patient.

A comparison of the direction and mean size of score changes under administration and review of verbal and gestural subtests for all subjects tested is presented in Table 7. It can be seen that score changes moved in both positive and negative directions for both the verbal and gestural modalities. However, on three out of the four

TABLE 7

A COMPARISON OF DIRECTION AND MEAN SIZE OF SCORE CHANGES FROM ADMINISTRATION TO REVIEW OF VERBAL AND GESTURAL SUBTESTS FOR TWELVE SUBJECTS

Modality	Subtest	Direction and Size of Change
Verbal	Subtest I	-.19
	Subtest IV	+.02
	Subtest IX	+.06
	Subtest XII	+.04
Gestural	Subtest II	-.23
	Subtest III	+.20
	Subtest V	-.02
	Subtest VI	-.03
	Subtest VII	-.04
	Subtest VIII	-.05
	Subtest X	-.03
	Subtest XI	-.04

verbal subtests (Subtest IV, Subtest IX, and Subtest XII) score changes were in a positive direction, while on seven out of the eight gestural subtests (Subtest II, Subtest V, Subtest VI, Subtest VII, Subtest VIII, Subtest X, and Subtest XI) score changes were in a negative direction. The mean size of score changes from the administration to the review, for the 12 subjects for each of the subtests, ranged from .02 of one point to .23 of one point. This is small if compared with the possible range of 15 points or the probable range of 10 points.

The direction and size of score changes under the conditions of administration and review were independent of the difficulty of the subtest. (See Appendix E.)

Types of Changes

Over all items and all tests given, 24.07 percent of the scores assigned during administration were changed during the review. From Figure 2 it can be seen that an 11 type response (a delayed incomplete) was altered most often--46 percent of the total number of times it occurred. The score that changed the least compared with the total number of times of occurrence was 15 (a complete response), which was altered 6 percent of the time.

In the examination of Table 8, which shows the distribution of changes over 120 items from the response level given during administration to that given during the review for the 12 subjects, it should be noted that the communicative ability of the particular 12 subjects tested may have skewed the results. However, it is of interest to note that the major type of score changes involved a decision related to delay (score changes of 15 to 13, 13 to 15, 12 to 11, and 11 to 12).

FIGURE 2

PERCENTAGE OF RESPONSE CHANGES FROM THAT ASSIGNED DURING ADMINISTRATION AT EACH OF THE LEVELS USED, OVER ALL ITEMS AND ALL TESTS GIVEN

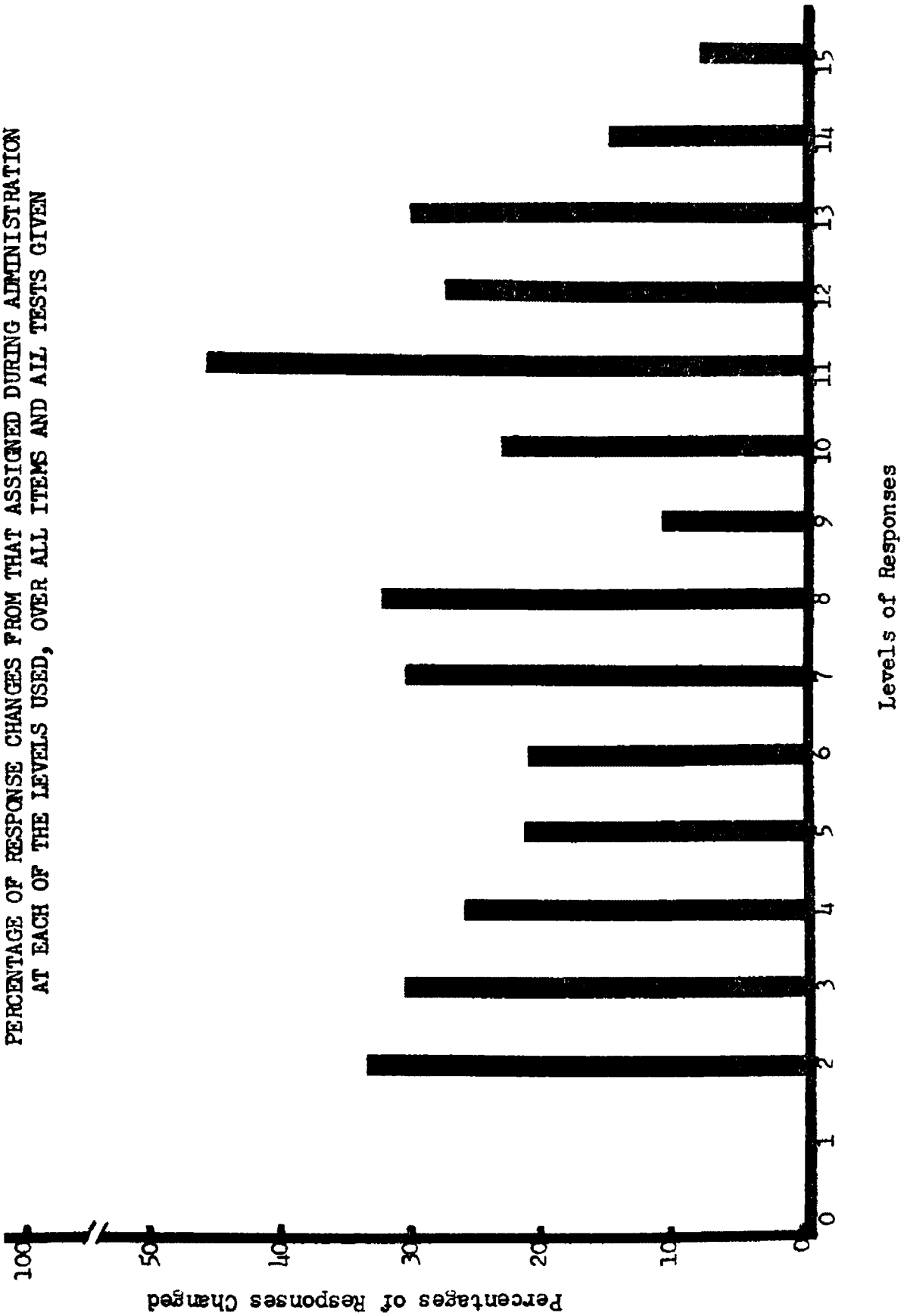


TABLE 8

DISTRIBUTION OF NUMBERS OF CHANGES OVER 120 ITEMS FROM THE RESPONSE LEVEL GIVEN DURING ADMINISTRATION TO THAT GIVEN DURING REVIEW FOR TWELVE SUBJECTS

Administration \ Review	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
16																	
15	6	20	6	18*	9	1	2			2							
14		8	122	4	4												
13		14*	7	119	3	4	2	2		2	1						
12		4	1	1	105	12*				8			1				
11			1	11	7*	57			1	6							
10				1			18			2			1				
9								46		1							
8								5	23	2							
7			1	1	3	4	1	3		74	1	4	3				
6		1	1	1						2	52	5		1			
5										5		58	6	1			
4								2		2		3	41	2	1		
3												9	8	62	1		
2												2		12	43		
1																	
0																	

* Change related to a decision about delay.
___ Number of times no changes in score occurred.

It also appears that the majority of score changes were approximately 1 or 2 points in magnitude except for 7 (related response) which changed 5 points 11 times, 4 points 10 times, with the maximum change of 8 points occurring twice. These 8-point changes were the largest that occurred under the two conditions in the study. In addition, score changes tended to cluster at the 10 type score and above, at the 6 type score and below, and throughout the 7 type score. However, as evidenced by the averages reported earlier, most of the large numbers of changes tended to cancel out each other. This can be seen from the relatively equal distribution of score changes on either side of the diagonal in Table 8.

Subjects and Clinicians

Four of the 12 subjects in the study were tested for the first time. A comparison of the number and size of score changes with this variable of whether the subject had previously been tested did not suggest any relationship. (See Appendix F.)

No apparent difference was found in consistency of scores under the conditions of administration and review and the amount of training or experience of the clinician. (See Appendix G.)

From the interview, one of the clinicians felt that the videotape review scoring was most accurate, one felt that the live presentation was most accurate, and the other four varied in their opinions depending on a particular patient or aspect of his behavior.

Only the emotional state of one of the 12 patients tested was judged by his clinician to have interfered with the administration and scoring of the test. (See Appendix H.)

CHAPTER IV

DISCUSSION

This study stemmed from the observation that in rescoring the PICA using videotape, the clinician often altered a number of assigned response scores. The aim, therefore, was to collect and analyze data to explore the possible causes, magnitude and consequences of these response score changes.

Analysis of Average Scores

Porch (1967, p. 37) compared the difference between subtest means and modality means of 30 patients for three scorers to establish inter-observer reliability. If the initial 12 subtests are taken, the average difference for these 12 subtests calculated from the difference between the highest and lowest of the three scorers was .17 points. The range of differences between the average scores for the 12 subtests under the administration and videotape review conditions, obtained in this study, was .18 points. Hence, using the average for 12 subtests, the change is very small and is almost equal to the findings in the inter-observer reliability studies conducted by Porch (1967).

A comparison of data from this study on the verbal and gestural subtests under the administration and review conditions, with the above mentioned reliability data (Porch, 1967), showed that for the verbal modality, the range between the highest and lowest differences (from

.03 to .50 points on this study) was slightly larger than that found by Porch (1967) (from .12 to .29 points). For the gestural modality, the range between the highest and lowest differences (from .01 to .36 points on this study) was slightly smaller than that found by Porch (1967) (from .05 to .61 points). However, the difference in modality scores under the two conditions could still be considered as minimal when the total possible range of difference (15 points) is noted.

Score Changes

The results of the average score change under administration and review were similar for each of the subtests, and were all less than .25 of one point. This was a smaller difference than that found by Porch (1967). The verbal subtests on the average in this study showed more and larger changes than did the gestural. This was contrary to the findings of Porch (1967), where scorer disagreement on the verbal items was found to be lower than on gestural items, as noted in the previous paragraph. It should be noted that in all of the above comparisons made between this study and the findings of Porch (1967), that this study was one of intra-observer reliability under two conditions (administration and review), while Porch (1967) investigated inter-observer reliability in direct observation of test administration. Thus, it may be argued that the comparisons were not valid. However, as no other criteria for the significance of the size of score changes was available, the results of Porch's (1967) reliability study were selected as a reference.

In spite of the small differences found between the average scores discussed earlier, over 24 percent of the time assigned scores

were changed from the administration to the review condition. Hence, these changes must tend to cancel each other out. That is, score changes tended to move in both a positive and negative direction and the magnitude of these changes were about equal in either of these directions over the VGS for each of the 12 subjects. Another factor was that the magnitude of changes was usually only one or two points.

Even though score changes tended to cancel each other out, it appeared that over all tests given, verbal subtest changes tended to move in a positive direction while gestural subtest changes tended to move in a negative direction. This would suggest that verbal responses appeared less impaired to the clinician on the videotape review while the gestural responses appeared more impaired. Verbal responses tend to be more discrete than gestural responses, e.g., to the question, "As completely as possible, tell me what you do with each of these," for a 15 (complete) score, the patient may answer, "I brush my teeth with a toothbrush." However, to the question, "As completely as possible, show me what you do with this," the patient must part his lips, exposing his teeth and demonstrate a brushing motion for a 15 score. Hence, because gestural responses are more fleeting and more complex, the clinician may have overestimated the patient's response during the live administration and this would tend to explain why gestural responses were scored lower on the videotape review. It is suggested by the writer that verbal responses were more easily recalled from the live presentation to the review because there are fewer of them (four verbal subtests to every eight gestural) and this may have been an aid in the second scoring, e.g., the scores became more intelligible.

Type of Response

An 11 response is described as "incomplete-delayed," that is, two dimensions of defective communication ability are involved. These require two decisions, one regarding promptness and one regarding completeness. None of the other PICA scores involves a double decision similar to this one, which may account for the relatively high number of the times the score changed (46 percent of the time). A complete, accurate, responsive, prompt and efficient score, a 15, was the most consistent score under the two conditions. This would suggest that clinicians have most difficulty judging impaired communicative ability, particularly if more than one dimension of judgment is involved.

One of the major reasons for change in scores between the two conditions appeared to be related to decisions concerning delays (see Table 8), but the caution regarding the communicative ability of the subjects in this study is repeated, that is, 10 of the 12 subjects functioned at an overall level of 10 or above and the judgment of delay is a major dimension in the assignment of score at this level, compared with subjects functioning below 10 where a delay in response is no longer recorded. The criteria for scoring delays vary according to the subtest. As an example, for subtests involving reading, the clinician was instructed to read the instructions over to himself twice and if the patient had not responded, a delay was scored; for subtest II, however, delay was defined as "any of the above responses, but delayed"; for subtest I, delay was defined as "response types that are significantly delayed, suggesting that the patient needs some additional processing time before he can respond" (Porch, 1973). During

the administration of the test, a delay in response can vary up to 30 seconds in duration, at which time a repeat of the instruction is given and another score assigned.

From the interviews with the clinicians involved, delay was judged according to a number of standards: (1) comparison with performance of a subject with normal communicative ability, (2) according to the general rate of response for that particular patient, or (3) a combination of (1) and (2). It is suggested that the large number of delay changes may be a factor of inadequate definition of a delay response. During the videotape review the clinicians appeared to be slightly more strict in their assessment of delay than during the administration. The writer also speculates that they became more consistent during the review when other pressures of test administration were not present.

From Table 8, the score 7 (related response) showed relatively larger changes under the two conditions than other scores and this suggests that clinicians may have difficulty differentiating between incomplete and related responses. Clinicians appeared to recognize related responses more easily under conditions of videotape observation. Again, it might be suggested that when the clinician is free of the pressure of test administration, he may be more sensitive to the patient's response. Another score that appeared to change under the two conditions was related to a decision on the intelligibility of scores (score changes between the response levels of 5, intelligible, and 2, attention). It is of interest to note that the trend was for these scores involving intelligibility to increase during the videotape

review, suggesting that the patients' responses became "more intelligible" and more differentiated during the review. Here again, the same hypothesis might be offered. When the clinician is free of test administration responsibility, he may be more sensitive to the meaning of the patient's distorted speech.

Clinicians

From the results obtained in this study, it would appear that following the basic training on the administration and scoring of the PICA as stated earlier, the amount of further training on this test has little relation to score inconsistencies under the two conditions of administration and review. These inconsistencies may be affected by such factors as level of communicative ability and the emotional state of the subject. However, it seems that the clinician's ability to obtain consistent average scores is maintained over 18 months after basic training. It is expected that clinicians involved in this study attempted to administer and score the PICA more accurately under the research conditions than they routinely did, but it is felt by the writer that this would not substantially change the results.

The clinicians offered the following comments when asked to compare the accuracy of assignment of scores under the two conditions:

- the videotape review allowed for more careful observations because it is possible to go back and verify them.
- the live presentation gave more sense of what the patient was doing because of the proximity.
- the live presentation was more accurate, particularly with a low level patient, because you were "testing in context."

- the videotape scoring was better because I did not have to control for my behavior such as facial cuing.

Conclusions

The results of this study suggest that the average scores on the PICA for VGS and for each of the verbal and gestural modalities, are stable under the two conditions of live presentation and videotape review. Hence, rescoring as a general practice is not suggested from these data. Clinicians could perhaps review scoring occasionally via videotape as a check on their continued reliability. It also appears that the use of videotape recordings in teaching the administration and scoring of the test is a valid procedure, because average scores for the live presentation and videotape review remained consistent. It should also be noted that verbal and gestural modalities remained consistent over the two conditions. This appears to be in agreement with past research which also indicates that the live presentation and the videotape review are highly similar.

However, 24 percent of the scores were changed from the administration to the videotape review; thus, 76 percent of the scores remained unchanged under the two conditions. The major reason for the score changes observed in this study include: (1) inadequate definition of delay, hence this response type tends to be subject to a variety of assignment criteria; (2) the relatively subjective response category of 7 (related) seems to be sensitive to repeated viewings via videotape, that is, the clinician may decide on one occasion that a response was "related", on another he may categorize the same response as more impaired and therefore an "error"; (3) clinician's difficulty

judging an 11 (incomplete-delayed) score which requires decisions on two dimensions of communicative impairment. It is suggested that when the clinician is free of the pressures of test administration during the videotape review, he may be more sensitive to the patient's responses, and therefore, score more accurately.

The present study employed a limited number of subjects and clinicians. It is suggested that follow-up studies to this present investigation involve a larger number of subjects, whose overall communicative ability is more representative of each of the levels of response. The majority of subjects in this study, on the average scores for VGS, functioned above the 10 response level. It is the writer's impression, based on the results of this study, that average scores for VGS and verbal and gestural modalities would always tend to be stable from the live administration to the videotape review, even if a larger and more representative sample of patients were used. Having this more representative sample of varying degrees of communicative involvement would allow for more definitive examination on the types of responses, e.g., delays, that are likely to change under the two conditions. It would also allow for investigation of the relationship between the patient's overall level of communicative functioning and the total number of score changes.

In addition, a panel of "experts" (persons demonstrating inter-observer reliability) could observe with the clinician under the two conditions so that a judgment of which condition is most accurate in terms of score assignment could be made. It is noted that these "experts" would be free of the pressures of test administration, and

therefore in a position to devote their entire attention to the scoring process. This would also allow for better interpretation of the finding which showed score changes in a positive direction for three out of the four verbal subtests and score changes in a negative direction for seven out of the eight gestural subtests.

CHAPTER V

SUMMARY

The aim of this study was to collect and analyze data in order to explore the possible causes, magnitude and consequences of assigned response scores on the verbal and gestural subtests of the PICA under two conditions. One of the conditions was live administration and scoring of the test, and the other a videotape review of the same test by the original administrator.

Six clinicians trained in the administration and scoring of the PICA and 12 subjects with varying degrees of communicative ability were used in the study.

Results showed that overall averages of VGS as well as the verbal and gestural modality averages made only slight changes under the two conditions, suggesting that the PICA is a reliable test. However, scores were altered under the two conditions over 24 percent of the time, so that these changes must tend to cancel each other out. The type of response that changed between the original administration to the rescoring was most often related to the 11 response (incomplete-delayed). It seems this score requires decisions on two dimensions of impaired communicative ability, where other scores require decisions about only one dimension at a time. Response types involving delay and those involving intelligibility were the other responses that

changed most frequently. It is suggested that a more definitive description of the category of delay is indicated. More careful selection of the patient population would be required, however, before this observation could be validated.

Implications and suggestions for further research were presented.

BIBLIOGRAPHY

BIBLIOGRAPHY

Boone, D. R., and Prescott, T. E. Speech and hearing therapy scoring manual: A manual for learning to self-score the events of therapy. University of Denver, Denver, Colorado, 1971.

Burkland, M. Use of television to study articulation problems. Journal of Speech and Hearing Disorders, 32, 1967, 80-81.

Croft, R. G., Stimpson, D. V., Ross, W. L., Bray, R. W., and Bieglic, V. J. Comparison of attitude changes elicited by live and videotape classroom presentations. Audio Visual Communication Review, 17, Fall 1969, 315-31.

Cullinan, W. L., and Prather, E. M. Reliability of 'live' ratings of the speech of stutterers. Perceptual and Motor Skills, 27(2), 1968, 403-409.

Denoff, M., Stenmark, D. E., and Smith, R. E. Comparison of videotape and face to face interviewing. Psychological Reports, 27(1), Aug. 1970, 53-54.

Flynn, Pauline T. Development and evaluation of videotape discrimination training programs. Unpublished dissertation, University of Kansas, 1970.

Kerlinger, Fred N. Foundations of Behavioral Research. Second edition. Holt, Rinehart, and Winston, Inc., 1973.

Norquist, V. M. A method for recording verbal behavior in free-play settings. Journal of Applied Behavioral Analysis, 4(4), 1971, 327-331.

Porch, B. E. Theory and development. Porch Index of Communicative Ability. Consulting Psychologists Press, Palo Alto.

Porch, B. E. Multidimensional scoring in aphasia testing. Journal of Speech and Hearing Research, 14(4), 1971, 776-792.

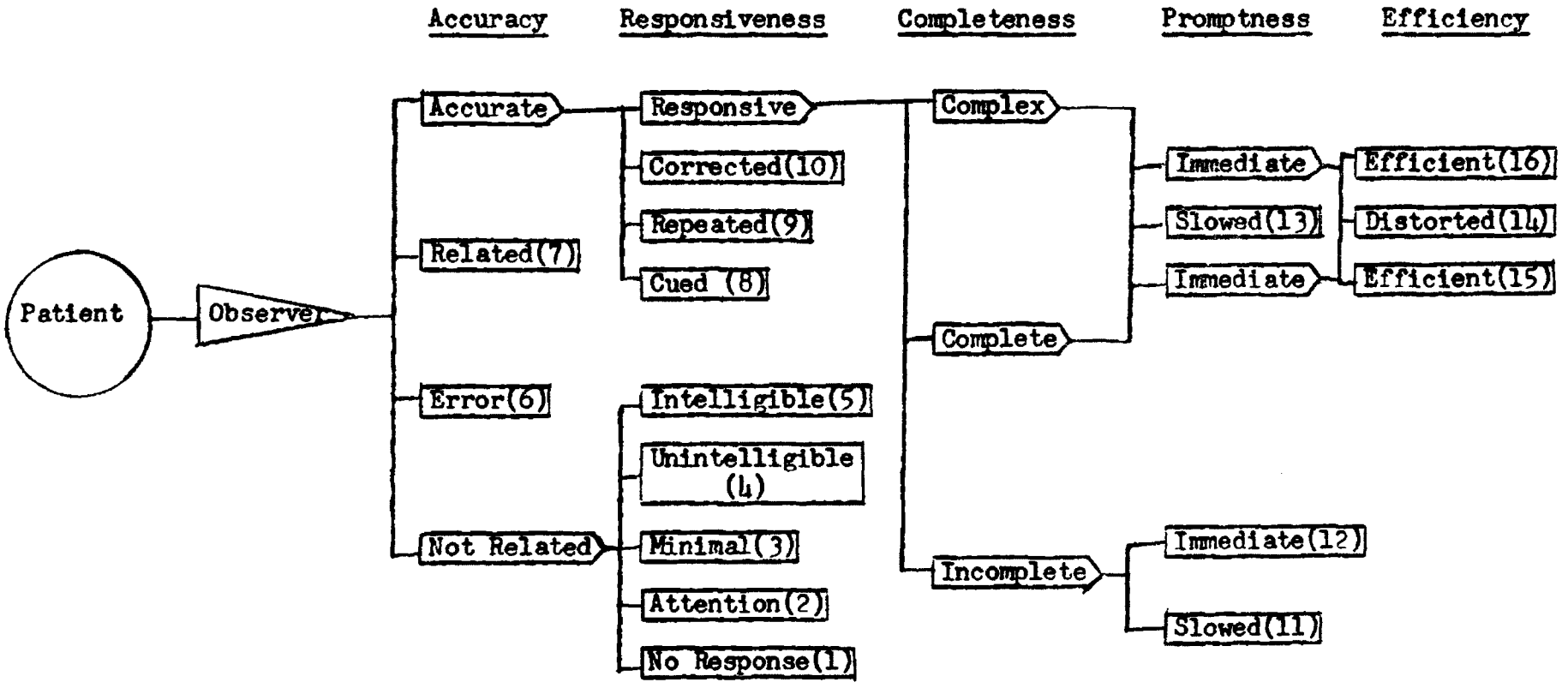
Porch, B. E. Administration, scoring, and interpretation. Porch Index of Communicative Abilities. Vol. 2, Revised Edition. Consulting Psychologists Press, Palo Alto, 1973.

Taylor, D. R., Lipscomb, E., and Rosemier, R. Live versus videotaped student-teacher interaction. Audio Visual Communication Review, 17, Spring 1969.

Wall, V. D. Jr., and Boyd, J. A. Channel variation and attitude change: live, written, and videotaped message. Journal of Communication, 21, 1971, 363.

APPENDIXES

APPENDIX A



BINARY CHOICE SCORING SYSTEMS BASED ON FIVE DIMENSIONS

Source: Porch Index of Communicative Ability - Theory and Development, Volume 1, p. 15, by Bruce E. Porch (1967).

APPENDIX B

QUESTIONNAIRE

1. Approximately how many times, within the past three months, have you administered the PICA?

   _____

2. What training did you have in administration and scoring the PICA?

   _____

3. How long has it been since you learned this test?

   _____

4. Have you tested either of these patients previously, using the PICA?

   Name _____ Yes / No

   Name _____ Yes / No

5. To what degree do you feel the patient's emotional state during the test (e.g., anxiety, frustration, anger) interfered with your administration and scoring of the test? (Mark the appropriate level according to the following scale).

Made scoring significantly more difficult				Did not make scoring more difficult
5	4	3	2	1

APPENDIX C

SIZE OF CHANGES CALCULATED FROM THE DIFFERENCE BETWEEN SUBTEST MEANS UNDER ADMINISTRATION AND REVIEW

Clinician	Subject	Subtest I	Subtest II	Subtest III	Subtest IV	Subtest V	Subtest VI	Subtest VII	Subtest VIII	Subtest IX	Subtest X	Subtest XI	Subtest XII
1	a	.2	-.2	.1	-.2	0.0	0.0	0.0	0.0	-.2	0.0	0.0	0.0
	b	.1	.2	.2	0.0	-.5	0.0	.2	0.0	.1	0.0	0.0	0.0
2	c	.8	-.4	-.2	-.1	0.0	.1	0.0	0.0	0.0	0.0	0.0	0.0
	d	-.9	.4	1.0	.1	-.4	0.0	-.1	0.0	.4	0.0	0.0	-.1
3	e	-1.1	0.0	.5	0.0	.2	0.0	.2	0.0	-.1	0.0	0.0	0.0
	f	-.4	.3	0.0	1.1	0.0	0.0	0.0	0.0	.2	0.0	0.0	-.5
4	g	-.1	-.9	-.1	0.0	.2	0.0	.1	0.0	-.1	0.0	0.0	.3
	h	.2	-.7	-.9	-.2	-.8	0.0	0.0	0.0	-.5	0.0	0.0	.7
5	i	0.0	.5	1.5	0.0	.7	-.6	-.8	0.0	.3	-.2	-.2	0.0
	j	.3	0.0	.1	-.5	0.0	.1	0.0	-.6	.9	-.4	-.3	.2
6	k	-1.4	-.4	-.4	1.3	-.5	0.0	0.0	0.0	.4	.2	0.0	-.1
	l	0.0	-1.0	.6	-1.3	.9	0.0	-.1	0.0	-.7	0.0	0.0	0.0

APPENDIX D

NUMBER OF CHANGES FOR EACH SUBTEST

Clinician	Subject	Subtest I	Subtest II	Subtest III	Subtest IV	Subtest V	Subtest VI	Subtest VII	Subtest VIII	Subtest IX	Subtest X	Subtest XI	Subtest XII
1	a	1	1	2	3	0	0	0	0	2	0	0	0
	b	2	1	1	0	4	0	1	0	1	0	0	0
2	c	6	2	3	1	0	1	0	0	0	0	0	0
	d	5	1	3	3	2	0	7	0	2	0	0	2
3	e	6	5	4	0	1	0	1	0	1	0	0	0
	f	1	4	0	5	2	0	0	0	2	0	0	6
4	g	0	3	3	2	2	0	1	0	1	0	0	0
	h	3	2	4	1	2	0	1	0	2	0	0	3
5	i	0	5	9	0	7	5	2	0	4	1	1	0
	j	6	5	5	5	0	2	2	3	7	2	1	2
6	k	5	5	2	4	4	0	2	0	2	1	0	0
	l	3	3	3	4	1	0	1	0	3	0	0	0

APPENDIX E

SUBTESTS RANKED IN ORDER OF DIFFICULTY COMPARED WITH THE DIRECTION AND MEAN SIZE OF SCORE CHANGES FOR ALL TWELVE SUBJECTS UNDER THE CONDITIONS OF ADMINISTRATION AND REVIEW

Subtest	Size and Direction of Score Change
I	-.19
II	-.23
V	-.02
IV	+.02
IX	+.06
III	+.20
VII	-.04
XII	+.04
VI	-.03
X	-.03
VIII	-.05
XI	-.04

The verbal and gestural subtests are ranked from most to least difficult.

APPENDIX F

RELATIONSHIP BETWEEN PREVIOUS TESTING OF THE SUBJECT AND THE NUMBER AND SIZE OF SCORE CHANGES BETWEEN ADMINISTRATION AND REVIEW

Previously Tested			Not Previously Tested		
Subject	Number of Changes	Size of Changes	Subject	Number of Changes	Size of Changes
a	10	.9	b	9	1.3
e	18	2.1	c	13	1.6
f	20	2.5	d	25	3.6
g	12	1.8	i	34	4.8
h	18	4.0	k	25	4.7
j	40	3.4			
l	18	4.6			

APPENDIX G

CLINICIANS RANKED IN ORDER OF THE AMOUNT OF EXPERIENCE COMPARED WITH THE SIZE AND NUMBER OF CHANGES MADE UNDER CONDITIONS OF ADMINISTRATION AND REVIEW

Clinicians Ranked (high - low)	Total Number of Changes for Both Tests	Total Size of Changes for Both Tests
3	38	4.6
1	19	2.2
4	30	5.8
5	74	8.2
6	43	9.3
2	38	5.0

APPENDIX H

DEGREE TO WHICH PATIENT'S EMOTIONAL STATE INTERFERED WITH ADMINISTRATION AND SCORING OF THE TEST

Made scoring significantly more difficult				Did not make scoring more difficult
5	4	3	2	1
	Subject j			Subject a
				Subject b
				Subject c
				Subject d
				Subject e
				Subject f
				Subject g
				Subject h
				Subject i
				Subject k
				Subject l