Temporal Reliability of the Feature Finders

Rosalie Carson

Eastern Illinois University

This research is a product of the graduate program in Communication Disorders and Sciences at Eastern Illinois University. Find out more about the program.

Recommended Citation

https://thekeep.eiu.edu/theses/3734
PAPER CERTIFICATE #2

TO: Graduate Degree Candidates who have written formal theses.

SUBJECT: Permission to reproduce theses.

The University Library is receiving a number of requests from other institutions asking permission to reproduce dissertations for inclusion in their library holdings. Although no copyright laws are involved, we feel that professional courtesy demands that permission be obtained from the author before we allow theses to be copied.

Please sign one of the following statements.

Booth Library of Eastern Illinois University has my permission to lend my thesis to a reputable college or university for the purpose of copying it for inclusion in that institution's library or research holdings.

2/20/73

Date

I respectfully request Booth Library of Eastern Illinois University not allow my thesis be reproduced because ________________________________

__________________________________________________________

__________________________________________________________

__________________________________________________________

__________________________________________________________

__________________________________________________________

Date Author
TEMPORAL RELIABILITY OF

THE FEATURE FINDERS

(title)

BY

ROSALIE CARSON

THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF SCIENCE

IN THE GRADUATE SCHOOL, EASTERN ILLINOIS UNIVERSITY

CHARLESTON, ILLINOIS

1973

YEAR

I HEREBY RECOMMEND THIS THESIS BE ACCEPTED AS FULFILLING

THIS PART OF THE GRADUATE DEGREE CITED ABOVE

DATE

2-20-73

DATE

Feb 19, 1973
ACKNOWLEDGEMENTS

The author wishes to extend a warm and hearty "thanks" to Dr. Lynn Miner for the unlimited amount of time he spent supervising her on this project. A special thanks is extended to the Head of the Department of Speech Pathology and Audiology, Dr. Thurman, for his assistance in supervising. Recognition is extended to Dr. Griffith for his advice and assisting from beginning to end. The author would like to express her gratitude to all of those who helped during the testing period. These acknowledgements cannot be ended without the author expressing her love and appreciation to her parents, Mr. and Mrs. Henry S. Carson, for their loving faith and assistance. Finally, the author wishes to express her utmost appreciative love to a dear and personal friend, Mr. Clarence Alexander Searcy, for his encouragement and understanding during the author's entire stay here.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>ii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF ILLUSTRATIONS</td>
<td>v</td>
</tr>
<tr>
<td>I INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>II REVIEW OF LITERATURE</td>
<td>7</td>
</tr>
<tr>
<td>III PROCEDURE</td>
<td>19</td>
</tr>
<tr>
<td>IV RESULTS AND DISCUSSION</td>
<td>24</td>
</tr>
<tr>
<td>V SUMMARY AND CONCLUSIONS</td>
<td>41</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>45</td>
</tr>
</tbody>
</table>


LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>List of Sources From Which Subjects Were Selected</td>
<td>19</td>
</tr>
<tr>
<td>2.</td>
<td>Chronological Age for Each Subsample</td>
<td>20</td>
</tr>
<tr>
<td>3.</td>
<td>Test-Retest Reliability Coefficients for Each Age Group Determined by the Pearson r</td>
<td>25</td>
</tr>
<tr>
<td>4.</td>
<td>Skewness and Kurtosis Values for Each Subsample</td>
<td>27</td>
</tr>
<tr>
<td>5.</td>
<td>Standard Error of the Mean for Each Subsample</td>
<td>34</td>
</tr>
<tr>
<td>6.</td>
<td>Standard Error of the Measure for Each Subsample</td>
<td>37</td>
</tr>
</tbody>
</table>


LIST OF ILLUSTRATIONS

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Graph Illustrating Mean Test-Retest Scores for Three-and-One-Half-Year-Olds</td>
<td>29</td>
</tr>
<tr>
<td>2.</td>
<td>Graph Illustrating Mean Test-Retest Scores for Four-Year-Olds</td>
<td>30</td>
</tr>
<tr>
<td>3.</td>
<td>Graph Illustrating Mean Test-Retest Scores for Four-and-One-Half-Year-Olds</td>
<td>31</td>
</tr>
</tbody>
</table>
CHAPTER I

INTRODUCTION

The primary purpose of this study was to investigate the temporal reliability of the Feature Finders. This chapter in particular, deals with the justification or rationale behind the study.

"Since a test must be reliable if it is to be valid, the statistical analysis of the quality of achievement tests scores ordinarily places primary emphasis on the reliability of those scores" (Ebel, 1965, p. 310). Five methods have been designated by Ebel for obtaining the independent measurements necessary for estimating the reliability of a test. They are: split half, Kuder-Richardson, test-retest, reader reliability and equivalent forms. Temporal reliability is thought to be one of the simplest methods of obtaining repeated measures for the same individuals with the same abilities. Giving the same test twice would yield a test-retest reliability coefficient (Ebel, 1965, p. 311). "A reliability coefficient obtained in this fashion is generally a fair indication of the test's reliability" (Lyman, 1964, p. 37).

Reliability data of the Feature Finders would provide important information regarding performance over a short period of time. The Feature
Finders is a global test of language. The test, designed by Griffith and Miner in 1969, measures a child's expressive as well as receptive language abilities with the environment being the primary source of linguistic content. According to Keenan, the test is based on the theoretical assumption "that if a child's attention is drawn to certain features that are relevant to solving a problem, he will attend to those features on future occasions" (Keenan, 1971). Gagne (Keenan, 1971) contends that when a child is faced with a problem where a particular feature is relevant, he will be able to transfer that which was learned from his previous experience which in turn should facilitate solving the new problem.

Keenan (1971) investigated the "clinical utility of the Feature Finders." Her basic procedures consisted of generating normative data on a population of preschool children within three designated age groups. She concluded the following in her study:

"1. Feature Finders is most discriminating for three-and-one-half-year-olds.

2. Four and four-and-one-half-year-olds rank order themselves in essentially the same manner. Three-and-one-half-year-olds do not.

3. The standard error of the mean indicates good reliability for all subsamples. Individual means would be expected to vary widely for three-and-one-half-year-olds, within a small range for four-year-olds, and moderately for four-and-one-half-year-olds.

4. Items decrease in difficulty as chronological age increases" (Keenan, op. cit.).
A replication of Keenan's pilot study on temporal reliability, using a larger number of subjects in each subsample, would provide additional information regarding the temporal reliability of the Feature Finders. According to Minifie, Darley and Sherman (1963), a variety of research and clinical tools have been designed by speech pathologists and others for describing and evaluating a child's language development. "Knowledge of whether these language abilities are variable or consistent within short periods would appear to be useful since both the researcher and the clinical worker need to know whether measures of language development may be considered fairly valid and reliable" (op. cit.).

Temporal reliability of seven language measures, including the mean length of response, was investigated in 1963 by Minifie, Darley and Sherman. One group of forty-eight five-and-one-half-year-olds and another group of forty-eight eight-year-olds, were selected on the basis of "age, sex, intelligence, socioeconomic status, hearing, physical status and language background." The obtained "intraclass correlation coefficient" ranged from .52 to .82. These researchers concluded that "language measures obtained from the 50-response language samples are not very consistent from day to day" (Minifie, Darley and Sherman, 1963). The MLR seems to be an unreliable measure of a child's expressive language. The test itself has been criticized by various researchers for several reasons. Many factors have been cited as possible causes for differences in the
recording and analysis of the MLR (Shriner, 1969). Shriner makes the point that examiners have to be particularly aware of those factors that can influence a child's responses. One other study on the temporal reliability of the MLR has been reported. In 1964, utilizing the "intraclass correlation coefficient," Fisher (Shriner, 1969) reported a reliability coefficient of .58 on the MLR and concluded that one cannot be sure that a "true" MLR exists from a 50-response language sample of any child.

Another measure of a child's expressive language has been brought to light since the MLR. Not only is one able to investigate a child's length of responses, but also the complexity of responses as well. The "length-complexity index" gives much more information about a child's morphological and syntactical rules manipulated in his language structure than does the MLR (Shriner, 1969). In 1969, Barlow and Miner investigated the temporal reliability of the LCI in comparison to the MLR. The resulting intraclass correlation coefficient was .80 which is much greater than that obtained for the MLR (.65). These researchers concluded that "as a language measure, the LCI is not as variable as the MLR; that it tends to measure children's language output more reliably over time" (op. cit.).

The results of the previous studies cited indicate the need for more reliable language measures. The addition of reliability data seems important to the clinical application of the Feature Finders. The data obtained by Keenan in studying the temporal reliability of the Feature Finders
in a pilot study, indicates the need for a more systematic replication using a larger sample size for each age group. According to Ebel, the reliability coefficient will typically be greater for scores from a large sample size having a wide range of abilities than from a group more homogeneous in abilities (Ebel, 1965, p. 336). Using five subjects from each age group, Keenan obtained reliability coefficients of \( .39 \) for three-and-one-half-year-olds, \( .90 \) for the four-year-old group and \( .80 \) for the four-and-one-half-year-old group" (Keenan, 1971). The method of sequential sampling was applied to the data. It was determined that "under 15 additional subjects" would be needed to acquire a reliability coefficient of \( .95 \) or greater for the two older groups (op. cit.).

**Statement of Purpose**

The primary purpose of this investigation was to evaluate the temporal reliability of the Feature Finders. The general methodological procedures consisted of testing and retesting three age groups of preschool children and analyzing the statistical results. Specifically, the following questions were posed at the outset of this study:

1. What is the temporal reliability of the Feature Finders at the age levels three-and-one-half, four, and four-and-one-half years?

2. What is the temporal reliability of the Feature Finders for each subtest at each age level?
3. What is the Standard Error of the Mean for each age level?

4. What is the Standard Error of the Measure at each age level?
CHAPTER II

REVIEW OF THE LITERATURE

The Feature Finders deals with learning. It focuses on attention and discrimination in addition to the role of language in the learning process. Although these terms have separate meaning, they are interrelated in the process of learning based on Feature Finders (Keenan, 1971).

Many approaches have emerged as bases for assessing a child's language. Most of the approaches are based on different theoretical assumptions. The purpose of this chapter is to review some of the theories on attention and discrimination upon which the Feature Finders is based.

The Feature Finders represents an awakening of interest in the role of attention in learning. The basic assumptions of the theory according to House and Zeaman are as follows:

"1. Attention is limited to only one of the many possible stimulus dimensions available to the subject at the moment of choice.

2. Subjects may learn to attend to or disregard stimulus aspects as a result of differential reinforcement.

3. Cues for instrumental learning are those aspects of the stimulus which are being attended to" (Keenan, 1971).
Data on research by House and Zeaman (Keenan, 1971), supports the hypothesis that the inability to attend to that which is relevant may be the cause of the learning deficit in retarded children.

Diller and Weinberg (1971) conducted a study on the "differential aspects of attention in brain-damaged persons." The subjects were 34 right hemiplegics and 35 left hemiplegics who were dominant on the right side, free from any acute medical or psychological problems, who were able to understand the demands of the tasks, and who were willing to participate in the study. Attention in these hemiplegics was assessed by presenting simple tasks that required a minimum amount of expressive language. The following measures were included in the battery:

1. Auditory cancellation of digits (ACD)
2. Auditory cancellation of letters (ACL)
3. Visual cancellation of digits (VCD)
4. Visual cancellation of letters (VCL)
5. Auditory digit span (ADS)
6. Visual digit span (VDS)
7. Auditory–visual integration task (AVI)

It was initially expected that a disturbance in attention to a specific modality would occur in the different types of aphasics. This was confirmed in the study. In regard to rehabilitation, several suggestions were made by Diller and Weinberg (1971). Their suggestions were as follows:
1. Train the left hemiplegic (LH) to attend to visual aspects of the environment to improve his performance in a variety of skills. Train the right hemiplegic (RH) to increase his span to improve his ability to process language.

2. Encourage the LH, who scans his environment too rapidly, to slow down. For the RH, who processes information too slowly, it would be helpful to speak to him at a slower rate.

3. For many habitual tasks, e.g., shoe lacing, the tasks demands are altered due to the condition of the disabled. Those tasks which represent something new requires the investment of attention (op. cit.).

Attention seems to be a very crucial aspect of the learning process in abnormals as well as normals. O'Connor (Fraser, 1972) has demonstrated that language deficits in imbeciles are largely due to inattention to relevant cues. Research by Suchman and Trabasso (Keenan, 1971) has lead them to the conclusion that the "ease or difficulty of learning seems to depend on the probability that a child would attend to and use the relevant stimulus feature."

The role of attention continues to be reviewed in recent literature. Posner and Boies (1971) discuss three main headings under which attention is relevant. The first is "alertness" in which they feel that attention is involved, especially when dealing with long tasks. The second heading is "selectivity" which requires the ability to choose relative information from one source rather than from another. Finally, a sense of attention is related to the notion of "a limited central processing capacity" (op. cit.).
The topic on how and when children begin to attend to relevant features and ignore the irrelevant features will now be discussed. Lehman (1972) conducted a study in which his primary objective was to investigate the issue on how and when children begin to attend to the relevant and ignore the irrelevant. Three experiments were conducted in which the nature of selective attention was explored by observing scanning strategies used by children in four grade levels; kindergarten, second, fourth and sixth. The strategies were observed on haptic matching tasks involving relevant and irrelevant tasks or redundant information. The subjects were required to match textured objects on shape or texture or match crosses on size. The results indicated developmental differences with the feature to be ignored and the extent to which attention was directed to the relevant feature. Two major questions arise from this study: Exactly where does the age change occur and How does it come about?

In answer to the first question, Lehman says that selective attention on some forms can appear very early depending upon the conditions involved. The five year olds, for example, focused their attention exclusively on the relevant features when they were clear to them. Performance varied somewhat with the irrelevant features. It appears that by the age of five, selective attention is well developed, therefore, the major age change would appear at an earlier age. No particular age is specified.
In answer to the second question, Lehman makes the following suggestions: First, review the literature by Munsinger and Kessen (1966), in which they conclude that children are less sensitive to redundancy than adults. Second, it is suggested that a follow-up study on the role of perceptual learning in different individuals at all grade levels be conducted. The results of Lehman's study showed that the younger children were just as economical as the older children in searching out features. This finding suggests that maybe intelligence does not play such a large role in attention as some believe. Finally, it is suggested that pertinent information may be provided if the relationship between selective attention and cultural experiences be explored. The overall results of Lehman's study lead him to the conclusion that selective attention appears to be a "multifaceted" rather than a "unitary" skill (op. cit.).

"Task orientation" versus "social orientation" was studied in young children in connection with their attention to relevant cues (Ruble and Nakamura, 1972). The purpose of Ruble and Nakamura's study was to examine how relevant cues given by an experimenter might affect the performance of task oriented versus socially oriented children on two games. The two games employed were the "squares (concept identification) game" and the "two object assembly tasks." The subjects were divided into "field dependents" and "field independents" on the basis of the Gerard rod and frame test. Field independents are those individuals who depend
totally on the task for making choice decisions. Field dependents are those individuals who also depend on the task but not as much as they depend on social cues such as the facial expressions of the experimenter. Because the field independents glanced more at the task and performed better than the field dependents, it was concluded that the glancing by the field independents was primarily a task oriented, information-seeking behavior rather than a socially oriented responding. The results indicated that the amount of glancing does not predict a child's ability to employ social cues but rather it depends on the direction in which the child is glancing and whether or not the cue given is relevant (op. cit.). Attention to relevant features and ignoring the irrelevant implies the concept of discrimination which seems inseparable to the role of attention in the learning process.

**Discrimination and Learning**

Discrimination, according to this investigator, can be thought of as a skill involving increasingly refined distinctions in choice decisions. Discrimination learning is thought to occur as far back as the infantile stage of development. According to Kephart (Keenan, 1971), an infant sees the world as a "vague mass characterized by qualities of the whole." Kephart further contends that as a child goes through various developmental stages, perceptual learning abilities are also developing, until he acquires
the perception of various distinctions (op. cit.). It has been suggested that children's discrimination learning in a variety of settings, often occurs under conditions of delayed reinforcements (Goldstein and Siegel, 1972). Goldstein and Siegel offer three explanations to explain how stimulus presence during delay facilitates discrimination learning. They are:

1. Observing responses made to the stimuli during may generalize to the prereseponse interval, resulting in a more careful comparative choice.

2. By providing the subject with a relevant focus for attention, stimulus presence during delay discourages extraneous or irrelevant competing attentional responses which retard discrimination learning.

3. Stimulus presence during delay provides additional opportunities for perceptual learning, that is differentiation of relevant stimulus features" (op. cit.).

In order to investigate children's discrimination learning during delayed reinforcement, Goldstein and Siegel conducted an experiment in which the primary purpose was to compare, directly, attentional versus perceptual learning functions of stimulus presence during delay. It was found that all delay conditions with additional stimulus exposure, both during delay or prior to response, was performed equally well, and the subjects' performance under those conditions equaled that of subjects in the immediate reinforcement conditions (Goldstein and Siegel, 1972).

The results of the study just cited implies an integration between perceptual learning and discrimination. Researchers such as Gibson
(Keenan, 1971) have described perceptual learning as "increasing differentiations of the features of the environment." Gagne (Keenan, 1971) describes two kinds of discriminations that must take place before learning occurs.

"1. We must learn to pay attention to only the relevant features of the stimuli involved and ignore the irrelevant.

2. We must then make the appropriate discriminations based upon those features."

Lawrenson and Bryant (1972) have designated two kinds of codes that may be involved in problem-solving tasks; the "relative" and the "absolute" codes. These researchers explain that when a child is instructed to discriminate between two stimuli that vary in size, for example, an absolute code would register only the actual sizes of the two stimuli. The relative code would be information that can be associated with one of the two stimuli, i.e., "that the larger of the two is correct."

Various sensory channels of communication are involved in discrimination learning. The point must be stressed that discrimination is not confined to the visual processes alone. The Feature Finders tests various environmental attributes that "are directly observable through three primary avenues of sensory communication: vision, audition, and taction" (Keenan, 1971).

McNinch and Richman (1972) studied the importance of auditory perceptual skills in discrimination, memory, integration and auditory visual
integration in a population of first graders. These researchers concluded that when word reading, vocabulary and paragraph meaning were the variables, reading in the first grade is more closely associated with auditory perceptual skills than to intellectual performance. The ability to hear is one aspect of the auditory channel that is vital, but of equal importance is "central processing" (Keenan, 1971). Chalfant (op. cit.) lists several tasks on which a child would perform poorly if his auditory processes are not intact. They are:

"1. Identification of the source of sound
2. Discrimination of sounds and words
3. Reproduction of pitch, rhythm and melody
4. Selection of significant from insignificant stimuli
5. Combination of speech sounds into words
6. Comprehension of the meaning of environmental sounds in general."

Performance on the Feature Finders, in general, is dependent upon a child's ability to process the verbal directives auditorially. "Temporal Relationships" has been included on the Feature Finders which requires the child's specific auditory attention to be able to reproduce various rhythmical patterns (Keenan, 1971). Berel, Diller and Orgel (1971) conducted a study employing a group of preschool children with cerebral palsy. These researchers used an electronically tuned xylophone in teaching various concepts such as numbers, names and labels. Berel and
others found that the ability to imitate short sound patterns was an attraction to brain-damaged children. It was concluded that auditory cues acted as facilitators for the sequential spatial activities that are required for imitating sound patterns.

Millar (1971) studied the importance of visual and haptic stimuli in recognition tasks. She found that visual cues were more effective than haptic cues. Even with a combined stimulation of visual and haptic cues, children's performance on visual recognition tasks was no better.

Yussen (1972) investigated his hypothesis that both visual and verbal experience will facilitate learning when relevant and interfere when they are irrelevant with later learning. Four year old preschoolers and seven year old second graders were used in Yussen's experiment. The subjects were trained to distinguish between two simultaneously presented forms under different conditions of visual and verbal highlighting of dimensions. All of the subjects were taught a transfer problem in which they had to reverse their earlier discrimination and, in addition, distinguish between two new forms. The results indicated that relevant verbal experiences facilitated learning only for the preschoolers and irrelevant verbal experience did not interfere with learning. It was also found that visual highlighting exerted no significant effects at both age levels. The visual channel has been tapped by several tasks on the Feature Finders that requires a child's visual attention. Most of the tasks require some verbal
involvement to some degree. Tasks which basically require verbal involvement includes some visual involvement also, especially in tasks which require a child to visually perceive an arrangement as the placement of a block in the box as in the "Spatial Relationships" subtest (Keenan, 1971).

Subtests as "Spatial Relationships, Color and Number" are designed to study the "visual-verbal relation" and are, therefore, "language bound" (op. cit.). According to Griffith and Miner, the term visual-verbal relations implies "a parallel system of specifying relationships or environmental arrangements" (op. cit.).

Chalfant (Keenan, 1971) discusses the importance of our senses of touch and body movement and concludes that these sensory channels provide us with such environmental information as:

1. Geometric information such as size and shape
2. Texture
3. Qualities of consistency such as hard or soft
4. Pain
5. Temperature and pressure"

There are several tasks on the Feature Finders that provide us with the above types of information. Because many decisions are made as a result of feedback from the sensory channels of touch and body movement, it is necessary for the individual to be able to attend to these.
In summarizing this chapter, the author would like to re-emphasize two basic theories which must be utilized for perceptual learning to take place; attention and discrimination. In the process of learning, attention to relevant features and ignoring the irrelevant features is a discriminative event in choice decisions. Transfer of learning has been demonstrated as being facilitated by attention and discrimination. A child's ability to attend to relevant features and ignore irrelevant features has been discussed as being dependent upon age. The Feature Finders is based on the theory that the environment is the most important source of linguistic content. Attention to various environmental attributes may facilitate problem-solving. Transfer of learning takes place when a child has previously solved a problem and is faced in another situation with a similar problem. Learning from a previous experience would, therefore, facilitate solving a new similar learning task.

Results of a Feature Finders test may be clinically useful to clinicians who are teaching language concepts such as names or labels. She would be able to determine from the results which environmental attributes are not being attended to.

As for the age factor which has been discussed as an important factor in a child's ability to attend and discriminate, this author feels that while age may be a factor, various accounts have yet to be offered on how children at very young ages learn to attend and discriminate as much as older children do.
CHAPTER III

PROCEDURE

Selection of Subjects.--Sixty subjects were selected on the basis of age from the available population of preschool children in the Central Illinois area. The following table lists the number of children selected from each source. Sources were selected on the basis of their willingness to cooperate.

TABLE 1

LIST OF SOURCES FROM WHICH SUBJECTS WERE SELECTED

<table>
<thead>
<tr>
<th>Source</th>
<th>Location</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>EIU Home Economics Nursery</td>
<td>Charleston</td>
<td>9</td>
</tr>
<tr>
<td>Mark Twain Elementary K</td>
<td>Charleston</td>
<td>12</td>
</tr>
<tr>
<td>Buzzard Laboratory School</td>
<td>Charleston</td>
<td>11</td>
</tr>
<tr>
<td>Methodist Day Care Center</td>
<td>Charleston</td>
<td>7</td>
</tr>
<tr>
<td>Tuscola Nursery Center</td>
<td>Tuscola</td>
<td>15</td>
</tr>
<tr>
<td>Other</td>
<td>Charleston</td>
<td>6</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td><strong>60</strong></td>
</tr>
</tbody>
</table>
The children were selected on the basis of availability. The total population of sixty subjects was divided into the following three age groups.

**TABLE 2**

**CHRONOLOGICAL AGE FOR EACH SUBSAMPLE**

<table>
<thead>
<tr>
<th>Age Range</th>
<th>M</th>
<th>SD</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>3, 3-3, 9</td>
<td>3.60</td>
<td>0.15</td>
<td>20</td>
</tr>
<tr>
<td>3, 9-4, 3</td>
<td>4.10</td>
<td>0.16</td>
<td>20</td>
</tr>
<tr>
<td>4, 3-4, 9</td>
<td>4.40</td>
<td>0.27</td>
<td>20</td>
</tr>
</tbody>
</table>

Selection of Examiners.—A team of ten examiners, nine females and one male, were selected from the Department of Speech Pathology and Audiology of Eastern Illinois University. All were trained in the area of speech pathology. All had taken one course in the area of child language and were enrolled in a second course. Although all of the examiners had assembled a Feature Finders kit, none was experienced in administering the test.

Multiple examiners were used in order that the resulting data would not be bound to the author. Examiners were assigned testing appointments on the basis of their availability.
Training of Examiners.--A training session was held in which the following points were reviewed:

1. Administration of the Feature Finders - The investigator demonstrated each task on the various subtests of the test along with the verbal directives. Each examiner performed each task after the investigator. The investigator then answered all questions asked about each task. On subtests which tests reception as well as expression, the examiners were told to only test receptive knowledge if an incorrect response was made on the expressive part of the task.

2. Recording Responses - Examiners were informed to mark correct responses with a plus and incorrect responses with a minus.

3. Scoring Responses - Questions concerning appropriate responses were discussed. The examiners were told to give credit for the receptive portion of a task if a correct response was made on the expressive part.

4. Test Forms - Each examiner was provided with a test form which contained the following information:
   a. Identification Outline
   b. Score Sheet
   c. Verbal Directives and Specification of Materials for each task.

The investigator pointed out various corrections that had to be made within the test forms. Each examiner was responsible for recording all necessary information called for on the test form and for scoring and totaling all responses for the children they tested.

Inter-Examiner Reliability.--Following the training session, the examiners watched a sample administration session being given on video tape. Each examiner was supplied with a test form to record responses
made and to compute a total test score. Using the total scores that each examiner recorded, a coefficient of agreement was calculated. A table was designed to tally the scores assigned by each examiner for each test item. These scores were compared with the resulting subtest scores of the investigator. Inspection of the table revealed one hundred percent agreement between the examiners and the investigator. Each examiner recorded the same score on each subtest as the investigator. The resulting reliability coefficient of agreement of 1.00 indicated very high reliability in the agreement between examiners.

**Testing Environment.**—Each child was tested individually in an environment with the least amount of distractions. Nearly all of the test situations afforded the examiner and child the privacy of a separate room with no interference. "Do not disturb" signs were posted on some of the testing rooms to avoid any interruptions.

**Method.**—A team of ten trained examiners, nine females and one male, administered the Feature Finders to preschool children within the designated age range of 3,3 to 4,9. After a maximum interval period of ten days with a minimum period of eight days, the same children were retested. No examiner retested a child whom she had previously tested. If the desired reliability coefficient of .90 or greater was not reached, then Silverman's (1968) method of sequential sampling was used to determine the number of additional subjects needed to obtain the desired coefficients.
Statistical Analysis.—The questions posed in this investigation were answered in the following manner:

1. What is the temporal reliability of the Feature Finders at the age levels three-and-one-half, four, and four-and-one-half years?

In order to determine the degree of correlation between the two sets of scores, the Pearson Product Moment Coefficient of Correlation was calculated. The resulting correlation coefficient indicated the test-retest reliability based on total test scores.

2. What is the temporal reliability of the Feature Finders for each subtest at each age level?

In order to compare results on individual subtests, three graphs were designed illustrating mean scores for each subtest on test and retest at each age level.

3. What is the standard error of the mean for each age level?

The standard error of the mean was computed to estimate the standard variance that would be found if the same population were retested.

4. What is the standard error of the measure at each age level?

The standard error of the measure was calculated to determine the amount of deviation in scores if individuals were retested.
CHAPTER IV

RESULTS AND DISCUSSION

The purpose of this investigation was to determine the temporal reliability of the Feature Finders. In order that meaningful interpretations be made, this chapter will report the statistical results of the data. Results will be discussed in relation to temporal reliability of the Feature Finders at age levels three and one-half, four and four and one-half. Temporal reliability of each subtest, standard error of the mean, and standard error of the measure will also be discussed for each age level.

**Temporal Reliability at Each Age Level**

To determine the temporal reliability of the Feature Finders at each age level, a Pearson product moment correlation coefficient was calculated for each set of scores. The resulting reliability coefficients are summarized in Table 3. Since reliability is a function of sample size (Horst, 1965, p. 337), reliability coefficients were expected to be substantially high with an N of twenty in each subsample. This N was chosen on the basis of the results Keenan obtained in her investigation of temporal reliability of the Feature Finders.
TABLE 3

TEST-RETEST RELIABILITY COEFFICIENTS FOR EACH AGE GROUP AS DETERMINED BY THE PEARSON r

<table>
<thead>
<tr>
<th>Age</th>
<th>r</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>3½</td>
<td>0.98</td>
<td>20</td>
</tr>
<tr>
<td>4</td>
<td>0.98</td>
<td>20</td>
</tr>
<tr>
<td>4½</td>
<td>0.99</td>
<td>20</td>
</tr>
</tbody>
</table>

Inspection of Table 3 reveals a high reliability coefficient for the three-and-one-half-year-olds. This indicates that the scores for this age group remained stable over a short period of time. The Pearson r for the four-year-olds was .98 indicating that these children did rank order themselves in the same manner on both trials of testing. Comparing the coefficient of .90, which Keenan obtained in her study for this age group with the coefficient obtained in this study, one can see that four-year-olds were relatively consistent in their rank ordering in both investigations. The obtained reliability coefficient for the four-and-one-half-year-olds was .99. Such a high coefficient was expected since Keenan's results revealed a coefficient that was substantially high, with an N of 5. The resulting reliability coefficient for the four-and-one-half-year-olds indicates that these individuals do rank order themselves in the same manner on both test and retest.
The shape of the distributions of scores was examined through measures of skewness and kurtosis. Skewness refers to a high concentration of scores at one end of the distribution curve and gradually flattening off at the opposite end. A positively skewed distribution is one in which the flat end of the curve contains the higher scores. A negatively skewed distribution occurs when the concentration of scores is at the higher end of the scale and tailing off at the lower end (Horst, 1966, p. 66). According to Griffin (1962), the formula used to determine the skewness in a distribution of scores suggests that an obtained value exceeding the range +0.50 indicates a considerable amount of skewness.

Kurtosis refers to the flatness or peakedness of a distribution, i.e., relative to a normal distribution. A distribution which is peaked is said to be leptokurtic and a distribution that appears flat is said to be platykurtic. Mesokurtic refers to a relatively normal distribution of measures (Horst, 1966, p. 67). The obtained values of skewness and kurtosis of this investigation will be compared with the values obtained by Keenan in 1971. The results are shown in Table 4.

For the three-and-one-half-year-olds, Keenan obtained a value representing a negatively skewed distribution and Carson obtained a value revealing a normal distribution. For the four-year-olds, Keenan obtained a value which resulted in a negatively skewed distribution, whereas Carson's obtained value reveals a normal distribution. For the
four-and-one-half-year-old group, Keenan's obtained value revealed a
distribution that was normal. Carson's obtained value reveals a positively
skewed distribution, i.e., on a scale, the flat end contains all the higher
scores of these children.

TABLE 4

<table>
<thead>
<tr>
<th>Age</th>
<th>Keenan Skewness</th>
<th>Carson Skewness</th>
<th>Keenan Kurtosis</th>
<th>Carson Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>3½</td>
<td>-0.77</td>
<td>-0.001</td>
<td>-9.41</td>
<td>0.46</td>
</tr>
<tr>
<td>4</td>
<td>-0.68</td>
<td>-0.28</td>
<td>-0.21</td>
<td>-0.01</td>
</tr>
<tr>
<td>4½</td>
<td>-0.22</td>
<td>0.51</td>
<td>-0.60</td>
<td>-0.34</td>
</tr>
</tbody>
</table>

For the three-and-one-half-year-olds, both Keenan and Carson ob-
tained values which reveal kurtosis to be mesokurtic, i.e., the individual
scores were normally distributed. In reference to the four-year-olds,
again both investigators obtained values which reveal a normal distribution
of scores. In regard to the four-and-one-half-year-olds, Keenan obtained
a value which indicated a platykurtic distribution. Carson's obtained
value for this subsample indicates that the scores were normally distributed.

In general, there were four differences between the Keenan and Carson
studies for comparisons of skewness and kurtosis. The differences found
may be the result of sampling error. Both investigators used subjects that were essentially from the same environment. The differences may also be due to examiner variability, or subject variability. One other factor must be considered and that is Keenan's N for each subsample was twice as large as the N this investigator used for each subsample. For the three-and-one-half-year-olds, Keenan used an N of 46, whereas this investigator used an N of 20. For the four-year-olds, Keenan employed an N of 41 and this investigator's N was again 20. For the four-and-one-half-year-olds, Keenan utilized an N of 47 and Carson's N was 20.

For further analysis of reliability, performance at the subtest level was examined. This was done by graphically plotting the mean test-retest scores of each subtest at each age level. These results are shown in Figures 1, 2, and 3 on pages 29, 30, and 31.

Review of the graph for the three-and-one-half-year-olds shows that the shape of the graph for both trials is relatively the same. In one instance the test-retest mean score was the same. This was on subtest number 7, "Visual Closure." The largest discrepancy between a test-retest mean score was .8. This was on subtest number 6, "Numeric Relationships." The discrepancies in the mean test-retest scores for each subtest did not necessarily show the discrepancy as an increase in the retest mean. Five subtests revealed higher means on the first testing.
FIGURE 1
MEAN TEST-RETEST SCORES FOR THREE-AND-ONE-HALF-YEAR-OLDS

Subtest Raw Score

Subtest Number

Test
Retest
FIGURE 2

MEAN TEST-RETEST SCORES FOR FOUR-YEAR-OLDs

Subtest Raw Score

1 2 3 4 5 6 7 8 9 10 11 12

Subtest Number

--- Test
--- Retest
FIGURE 3

MEAN TEST-RETEST SCORES FOR FOUR-AND-ONE-HALF-YEAR-OLDS
Various factors may have accounted for this, such as subject variability, examiner variability or no learning effect had taken place during the interval period. There was evidently some learning taking place on other subtests. The greatest variance in performance is shown on subtest number 10, "Temporal Relationships." This subtest was considered by the investigator to be a very difficult task to perform for the three-and-one-half-year-olds.

Inspection of the graphs for four-year-olds reveals relatively consistent test-retest mean scores. The largest discrepancy between test and retest mean is at subtest number 1, "Spatial Relationships." The retest mean on this subtest was .6 more than the test mean. This probably represents chance variation. One striking coincidence is that the same subtests that show the largest discrepancy for the three-and-one-half-year-olds, reveals discrepancies for the four-year-olds. Although the discrepancies are not more than .3 between test and retest, the discrepancies are at subtest number 6, "Numeric Relationships" and subtest number 10, "Temporal Relationships." Essentially, the mean test-retest shapes of the graphs are the same for the four-year-olds.

The graphs for the four-and-one-half-year-olds again reveals mean test-retest shapes to be essentially the same. For this age group, the largest discrepancy is surprisingly at subtest number 2, "Gross Color." This is probably due to the individual's attention being focused on other
dimensions of the stimuli rather than the obvious relevant features.

In general, there are no discrepancies greater than .8. The average amount of discrepancy on each subtest for each subsample was .2, which is considered quite small. Keenan's largest discrepancy for all subsamples was 1.4. Also, in this investigation, discrepancies did not necessarily show up as increases in retest means. For all subsamples, about half of the subtests' test means were greater than the retest means.

**Standard Error of the Mean**

Thus far, temporal reliability of the Feature Finders has been analyzed in regard to descriptive methods. Descriptive methods only provide a summary of an obtained set of scores. What is important now is how much confidence can be placed on descriptive statistics. Estimates of error are means by which statistical inferences can be made (Lyman, 1963, p.59).

The standard error of the mean reveals how much variance would exist in the means of one sample compared to the means of another sample if they were retested. The standard error of the mean was computed for each subsample and illustrated in Table 5.

The mean for the three-and-one-half-year-olds is 48.8, with a standard deviation of 7.1 and the standard error of the mean is 1.62. This is interpreted to mean that with 68 percent confidence, it can be determined that the sample mean for this subsample will not vary from the population mean more than ±1.62.
TABLE 5
MEANS, STANDARD DEVIATIONS AND STANDARD ERROR OF THE MEAN FOR EACH SUBSAMPLE

<table>
<thead>
<tr>
<th>Age</th>
<th>M</th>
<th>SD</th>
<th>SEx</th>
</tr>
</thead>
<tbody>
<tr>
<td>3½</td>
<td>48.9</td>
<td>7.1</td>
<td>1.62</td>
</tr>
<tr>
<td>4</td>
<td>54.9</td>
<td>8.1</td>
<td>1.85</td>
</tr>
<tr>
<td>4½</td>
<td>57.1</td>
<td>5.5</td>
<td>1.27</td>
</tr>
</tbody>
</table>

The mean for the four-year-olds is 54.9, with a standard deviation of 8.1 and the standard error of the mean is 1.85. With 68 percent confidence, the mean is likely to vary ±1.85 from one sample to the next 68 percent of the time.

The mean for the four-and-one-half-year-olds is 57.1, with a standard deviation of 5.5 and the standard error of the mean is 1.27. This result indicates that with 68 percent confidence, the population mean is likely to vary ±1.27 from one sample to the next 68 percent of the time.

Based on the high reliability coefficients that were obtained for each subsample, the standard error of the means were not expected to vary to a large degree from the population means. The largest variance was ±1.8.
In analyzing the significance of difference of the means for each subsample obtained by Keenan and Carson, the standard error of difference between the means was computed (Kerlinger, 1963, p. 178). The resulting standard error of difference for three-and-one-half-year-olds was 3.26. The standard error of difference for four-year-olds was 2.89 and the standard error of difference for four-and-one-half-year-olds was 3.21. All were significant beyond the .01 level of confidence, i.e., the means from both investigations of each age group were significantly different from each other with only one percent chance that they are not. These differences may be accounted for due to the fact that Keenan utilized means for analyzing central tendency of a group of individual scores that were negatively skewed. In this case, the mean would tend to be exaggerated. Medians would have been a more acceptable way of describing central tendency of a skewed distribution. Observation of the medians for each subsample of both investigations, reveals a very close relationship. This was further investigated by computing the standard error of difference between the medians of each age group (Garrett and Woodworth, 1954, p. 232). The resulting standard error of difference for three-and-one-half-year-olds was 1.68. The standard error of difference for four-year-olds was 1.84 and the standard error of difference for four-and-one-half-year-olds was 6.3. The results of the differences in the medians for the two younger groups was non-significant at the .05 level.
of confidence, i.e., the medians from Keenan and Carson's study of the two younger age groups were so close that there is only five chances out of a hundred that they are significant. For the four-and-one-half-year-olds, the result was significant beyond the .01 level, indicating a significant difference between the medians. This can be accounted for by comparing the results of skewness for the above age group from the two investigations. Keenan's obtained skewness value indicated a normal distribution whereas Carson's obtained value of skewness indicated a positively skewed distribution.

From a more logical viewpoint, this investigator feels that her values of skewness and kurtosis would be more acceptable than Keenan's since in essentially all instances the distribution shapes were normal.

The above results seem to lead into another conclusion. Since median values of Keenan's investigation were closer to the median values of Carson's study, this implies the possibility that maybe percentile ranking is a more reliable measure for describing the position of scores. This has yet to be investigated. Lyman contends that "percentile rank is a better position indicator because it makes allowances for differences in group size" (Lyman, 1966, p. 47). It was emphasized earlier that Keenan's N for each of her subsamples was twice as large as the N of Carson's subgroups.
Standard Error of the Measure

The standard error of the measure reveals deviations that would be found if individuals were retested. This measure gives an estimation of how much individual scores would vary from his true score. The following table illustrates the standard error of the measure computed for each subsample.

<table>
<thead>
<tr>
<th>Age</th>
<th>M</th>
<th>SD</th>
<th>SEmeas</th>
</tr>
</thead>
<tbody>
<tr>
<td>3½</td>
<td>48.9</td>
<td>7.1</td>
<td>1.00</td>
</tr>
<tr>
<td>4</td>
<td>54.9</td>
<td>8.1</td>
<td>1.14</td>
</tr>
<tr>
<td>4½</td>
<td>57.1</td>
<td>5.5</td>
<td>0.55</td>
</tr>
</tbody>
</table>

"The smaller the standard error of the measure, the more reliable the test and the more confidence we can place in any score obtained by using a test" (Downie and Heath, 1959, p. 222). The resulting standard error of the measure for three-and-one-half-year-olds is 1.00. This indicates that 68 percent of the time, an individual's true score would lie ± 1.00 between his obtained score and scores. In considering the obtained reliability coefficient for this age group (.98), standard error of the measure was not expected to be much larger than that obtained.
The standard error of the measure for the four-year-olds is 1.14. This value indicates that a child's obtained score would not vary more than ± 1.14 from his true score 68 percent of the time. Although the standard error of the measure for this subsample is greater than that obtained for the three-and-one-half-year-olds, it still indicated a substantial amount of reliability which can be placed on individual scores obtained.

The standard error of the measure for the four-and-one-half-year-olds is 0.55, which is small compared to the other two groups. The obtained value for this subgroup indicates that an individual's score would not vary more than ± 0.55 from his true score 68 percent of the time. Since the reliability coefficient for this subgroup of .99 was greater than that obtained for the other two subgroups, the standard error of the measure for four-and-one-half-year-olds was expected to be smaller than that obtained for the other two groups.

In general, standard error of the measure revealed that individual scores did not vary more than ± 1.14 from his true score. These results indicate that much reliability can be placed on individual scores as well as the set of scores for a subsample.

The results of the present study as reported in this chapter can be summarized as follows.
Reliability of the Feature Finders in terms of individuals' ability to rank order themselves in the same manner upon retest was examined. Sixty subjects ranging in age from three-and-one-half to four-and-one-half years were tested with the Feature Finders. The resulting twenty subjects of the three age groups were retested after a maximum interval period of ten days with a minimum of eight. A Pearson $r$ was computed for each set of scores for each subsample. The obtained coefficient for the three-and-one-half-year-olds was a substantially high 0.98 which indicated that these individuals did rank order themselves in a similar manner upon retest. The resultant coefficient for four-year-olds was 0.98 which again indicated a substantial amount of test-retest agreement. The coefficient for four-and-one-half-year-olds was 0.99. This very high coefficient reveals that these individuals did rank order themselves in essentially the same manner.

The distribution of total test scores were analyzed in terms of variability and shape. Measures of skewness and kurtosis were utilized. The obtained values of skewness indicated a normal distribution of scores for three-and-one-half-year-olds and four-year-olds. The value obtained for four-and-one-half-year-olds indicated a positively skewed distribution. The discrimination power is located at the higher end of the range.

Kurtosis values for all age groups indicated a mesokurtic distribution, i.e., the scores of all subsamples were normally distributed.
Graphs were plotted to further analyze temporal reliability of the Feature Finders at the subtest level. Mean test-retest scores for each subtest on each subsample were plotted. In general, mean subtest scores on the trials were very close. Discrepancies which existed were very small and did not necessarily indicate an increase upon retest.

The standard error of the mean and the standard error of the measure were calculated to determine the stability of the measures. The standard error of the mean for each subsample did not vary more than ±1.85 from the population mean. The standard error of the measure for three-and-one-half-year-olds was 1.00 which indicated that an individual's true score would vary somewhere between 1.00 ± his obtained scores. The resultant standard error of the measure for four-year-olds was 1.14. This indicated a substantially small degree of variance between individual obtained scores and true scores. The standard error of the measure for four-and-one-half-year-olds was 0.55 which indicated a very small degree of variance between obtained scores and true scores.

In general, all subsamples acquired very small standard errors of the measure which was initially anticipated after acquiring very high reliability coefficients.
CHAPTER V

SUMMARY AND CONCLUSIONS

The purpose of the present study was to investigate the temporal reliability of the Feature Finders. The basic procedure consisted of testing and retesting a population of preschool children and analyzing the statistical results. Specifically, the following questions were posed at the onset of this investigation.

1. What is the temporal reliability of the Feature Finders at the age levels three-and-one-half, four, and four-and-one-half years?

2. What is the temporal reliability of the Feature Finders for each subtest at each age level?

3. What is the standard error of the mean for each age level?

4. What is the standard error of the measure at each age level?

The Feature Finders was administered to sixty preschool children ranging in age from three-and-one-half-years to four-and-one-half-years. The subjects were divided into subgroups according to age at six month intervals. The resulting N for each subsample was twenty. For test-retest data, all of the children from each subsample were readministered
the Feature Finders after a maximum interval period of ten days, with a minimum interval period of eight days.

The Pearson product moment correlation coefficient was computed to determine the temporal reliability for each age subsample. Temporal reliability of the Feature Finders for each subtest was analyzed by means of a graphical representation of the test-retest mean scores of each subsample. The distribution of Feature Finders scores was described in terms of skewness and kurtosis. Standard error of the mean and standard error of the measure were also computed and discussed.

Conclusions

The statistical analyses resulted in the following conclusions concerning the temporal reliability of the Feature Finders test.

1. The temporal reliability of the Feature Finders
   a. was .98 for three-and-one-half and four-year-olds.
   b. indicated that the above two subgroups did rank order themselves in essentially the same manner.
   c. was .99 for four-and-one-half-year-olds.
   d. indicated that the above subgroup did rank order themselves in the same manner.

2. The distribution of mean Feature Finders scores
   a. was mesokurtic and indicated a normal distribution of scores for three-and-one-half-year-olds.
b. was mesokurtic and indicated a normal distribution of scores for four-year-olds.

c. was positively skewed for four-and-one-half-year-olds.

d. did not discriminate those individuals whose attention to features was poor.

e. was mesokurtic for four-and-one-half-year-olds.

3. Standard error of the mean

a. was within 1.85 for all subgroups.

b. indicated that a sample mean did not vary much from the population mean for all subsamples.

c. further indicated high reliability of the Feature Finders for each subsample.

4. Standard error of the measure

a. did not vary more than 1.14 for all subsamples.

b. indicated a small amount of variance between individuals' scores obtained and their true score.

c. was interpreted to mean that a high amount of reliability can be placed on each individual's scores as well as the set of scores.

In general, temporal reliability of the Feature Finders for each age level was very high. The investigator would like to insert that the results of this study reveal also the clinical usefulness of the Feature Finders in terms of the power to discriminate those individuals whose attention to relevant features is poor.

In light of the findings regarding the significant differences between
Keenan's and Carson's means for each subsample, and the insignificant differences found between their medians, further investigation seems implicated. New light would be added to the world of statistics if it is discovered that percentile ranking is a more reliable indicator of scores' position in comparison to the mean as a measure of central tendency. Evidence which gives support to the need for further investigation of the above has been reported in this study.

Now that favorable reliability data is gathered on the Feature Finders, the next likely course to take would be the investigation of test validity. The author would like to emphasize that high reliability data does not necessarily mean that a test is valid. High validity data on the Feature Finders would no doubt be a great asset to the repertoire of global language measures.
BIBLIOGRAPHY

Books


Articles


45


McNinch, G. and Richman, "Auditory Perceptual Tasks as Predictors of First Grade Reading Success," Perceptual and Motor Skills, 35: 1, 1972.


