A CONCEPTUAL SYSTEM FOR IDENTIFYING TEACHER BEHAVIORS IN PHYSICAL EDUCATION ACTIVITY CLASSES

by

Judith C. Showers

A Dissertation Submitted to
the Faculty of the Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Doctor of Education

Greensboro
1974

Approved by

Gail M. Dennis
Dissertation Adviser

APPROVAL PAGE

This dissertation has been approved by the following committee of the Faculty of the Graduate School at The University of North Carolina at Greensboro.

Dissertation Adviser Gail M. Hennis

Oral Examination Committee Members Pauline C. Leffler

Elizabeth C. Hansford

Rosemary McGee

Wallace B. Phillips

August 24 1973
Date of Examination

SHOWERS, JUDITH C. A Conceptual System for Identifying Teacher Behaviors in Physical Education Activity Classes. (1974) Directed by: Dr. Gail Hennis. Pp. 94

The present study was designed to develop a scale which may be used to observe and describe the frequency of occurrence of selected teacher behaviors in physical education activity classes. Related problems were to study the objectivity and validity of the scale constructed.

Seven concepts generally recognized in the literature as reflecting effective teaching were chosen for the study, and 31 behavioral correlates of these concepts were developed through documentary analysis and introspection. These behavioral items were submitted to a jury of nine experts for validation. In a pilot project, three observers used the scale to rate five teachers, and completed three daily ratings and a final composite form for each instructor. During the observations, judges rated each of the items as occurring frequently or always, sometimes, or seldom or never. The composite rating was completed without reference to the first three ratings. The objectivity for a total of 60 independent ratings was .73. The operational validity of the trial scale was discussed with the observers. Data from the jury members and the raters indicated that the scale was valid, and provided information which resulted in six minor revisions in the trial scale. The final scale included four broad categories (clarity and knowledge of subject, friendliness and interest in students, enthusiasm and sense of humor, and fairness) and 35 behavioral items.

The procedure used for studying the objectivity of the trial scale was utilized again in the actual study. Three judges observed five teachers in various physical education activity settings (archery, body mechanics, beginning swimming, tennis, and track and field) and completed three daily ratings and a composite rating for each. The data from 60 paired observations were analyzed by using the Pearson product-moment method of correlation, Fisher's z method for averaging correlation coefficients, and the percentage of agreement method.

The objectivity coefficients ranged from .63 to .84, with an overall coefficient of .76. Agreement was highest when observing teachers in tennis and swimming classes, and lowest when rating teacher behaviors in an archery class. The findings were significant at the .01 level, and the hypothesis that no significant relationship existed between or among the results obtained by independent scorers using the scale under the same circumstances was rejected.

An analysis of the objectivity on each of the items revealed a range of agreement of 37 to 97 per cent. At least two observers agreed 70 per cent of the time or more on the frequency of occurrence of 32 of the 35 items.

The operational validity of the scale was verified by the observers at the conclusion of the observations, and the findings indicated that: 50 minutes was sufficient time for marking the scale during an observation; three observations were sufficient prior to completing a composite rating; and

various types of activity settings may affect the objectivity and operational effectiveness of the scale differently.

ACKNOWLEDGEMENTS

The writer wishes to express appreciation to her adviser, Dr. Gail Hennis, for her patience, encouragement, and continued assistance throughout the preparation of this paper.

Gratitude is also expressed to the Department of Health, Physical Education, and Recreation at Oklahoma State University for allowing this study to be conducted. Special appreciation is extended to the observers and teachers who participated in the investigation.

TABLE OF CONTENTS

Chapter Page

LIST OF TABLES

CHAPTER I

INTRODUCTION

Skill and excellence in teaching are of primary concern to administrators, teachers and students in higher education. Although numerous studies have been conducted on desirable teacher characteristics and teacher effectiveness, few facts have been established (2, 14, 64). There is no doubt that increased understanding of teacher behavior and teacher effectiveness would benefit students, teachers, teacher educators, and administrators alike; therefore, "research toward its understanding must continue" (2:vi).

NEED FOR THE STUDY

There seems to be an increasing movement away from the application of uniform criteria to the observation and evaluation of teachers in all situations (10, 33, 46). Educators generally agree that more information is needed about the characteristics, skills, and behaviors of effective teachers in specific situations and cultural settings, since the importance of a particular criterion of teaching effectiveness may vary from student to student and from class to class (2, 10, 16, 46, 56, 64, 66).

Several reasons have been given for studying teaching effectiveness. Identification of desirable characteristics,

skills, and behaviors may provide a basis for the improvement of teacher preparation programs, teacher selection procedures and teaching quality, and for administrative decisions on academic rank, tenure, salary, and merit raises (2, 8, 22, 28, 40, 69).

A review of the related literature reveals that, although a proliferation of research exists regarding characteristics of effective teachers, more studies are needed to determine specific skills and behaviors which are correlates of these characteristics (11, 17, 49, 57). Concepts such as enthusiasm, clarity, fairness, interest, friendliness, sense of humor, and knowledge of subject need to be operationally defined (17). The writer, therefore, became interested in studying specific behavioral correlates (defined as operational definitions) of characteristics which have been identified as showing strong relationships to teaching effectiveness, and developing a conceptual system for identifying teaching behaviors in college physical education activity classes.

STATEMENT OF THE PROBLEM

The present study was designed to develop a scale which may be used to observe and describe the frequency of occurrence of selected teacher behaviors in physical education activity classes. The behaviors chosen for inclusion on the scale were decided upon after a review of the literature, introspection, and validation by a jury of experts. A related problem was to study the operational validity of the items on the scale. A

second related problem was to determine the objectivity of utilizing the scale constructed.

The three observers who used the scale during the study were graduate students at Oklahoma State University. The five teachers observed during the study were members of the teaching faculty in the Department of Health, Physical Education and Recreation at Oklahoma State University during the spring semester, 1973. Data were collected by means of rating scales and interviews.

PURPOSE OF THE STUDY

The general purpose of the present study was to develop an instrument which may be used to observe and describe the frequency of occurrence of selected teacher behaviors in a physical education activity setting. Specific purposes included the following:

1. To identify characteristics of effective teachers.
2. To identify behaviors which are correlates of the characteristics of effective college teachers.
3. To construct a scale which may be used to observe and describe the frequency of occurrence of selected teacher behaviors in college physical education activity classes.
4. To establish the validity of the scale constructed during this study.

5. To study the objectivity of rating the frequency of occurrence of the teacher behaviors selected for use on the scale constructed in this study.

Hypothesis

The hypothesis tested in this study was based on the assumption that there were identifiable characteristics recognized as describing the effective college teacher of physical education. It was further assumed that there were identifiable behaviors which were correlates of the characteristics of the effective college teacher of physical education. The following hypothesis was tested:

Regarding the use of the scale constructed in this study, there is no significant relationship between or among the results obtained by independent scorers using the scale under the same circumstances. For all of the observer comparisons, the 5 per cent level was selected for rejection of the null hypothesis.

DELIMITATIONS

The following delimitations describe the scope of the present study:

1. The members of the jury of experts were selected by the investigator.
2. The teachers and classes observed were not randomly selected.

3. The five teachers observed were members of the teaching faculty in the Department of Health, Physical Education and Recreation at Oklahoma State University, Stillwater, Oklahoma.
4. The three observers were graduate students in the Department of Health, Physical Education and Recreation at Oklahoma State University, Stillwater, Oklahoma.
5. The classes in which the observations were made were archery, body mechanics, swimming, tennis, and track and field.
6. Four of the classes met three times a week for 50 minutes; one class met twice a week for 75 minutes. All classes were observed during three consecutive instructional class sessions within a three-week period of time.

CHAPTER II

REVIEW OF SELECTED RESEARCH

A survey of the related literature revealed that numerous studies have been conducted in which students and teachers have described characteristics which they consider desirable in effective teachers. In addition, a number of attempts have been made to analyze teacher behavior and classroom climate in order to determine what constitutes effective teaching.

CHARACTERISTICS OF EFFECTIVE TEACHERS

Perceptions of Students

In an early investigation conducted by Davis and reported by Beecher (1), over 13,000 students were asked to list the qualities characterizing their best teachers. Qualities noted most frequently were: knowledge of subject matter, good character, fairness, sense of humor, discipline, ability to hold interest, clearness, willingness to help, personality, sociability, patience, and appearance. A survey of 320 students by MacDonald (37) revealed results similar to the Davis study. Fairness, friendliness, appearance, knowledge of subject, and sense of humor were again mentioned most frequently. Other qualities listed among the most important were sincerity, interest in individual students, sympathy, and a good voice.

Results of studies by Schaffle (53) and Tiedeman (1) indicate that junior high and high school students agree to some extent on characteristics of outstanding teachers. On a questionnaire distributed by Schaffle to 820 ninth-grade pupils and 850 twelfth-grade students, both groups listed impartiality, clarity in explanation, knowledge of subject, interest in students, ability to get the subject across, and pleasant disposition among the top ten. Eight thousand opinions of junior high pupils collected by Tiedeman revealed that they, too, considered clarity and fairness, in addition to friendliness and sense of humor, important prerequisites to effective teaching.

The concepts of fairness and clarity appeared again in a summary by Hart, cited by Beecher (1), of the opinions of 10,000 students regarding desirable teacher behavior patterns. Hart's study also reported cheerfulness, friendliness, enthusiasm, and knowledge of subject as important qualities of effective teaching. A survey by McComas (40) revealed identical results.

Langen (35) reviewed the literature and selected the 41 most frequently mentioned characteristics of a good college teacher. He then conducted an investigation to assess the relationship of each item to students' judgment of teaching effectiveness. The behaviors selected most frequently by students as being important to the effective teacher were: interprets abstract ideas and theories clearly, gets me interested in the subject, increases my skills in thinking, helps broaden my interests, stresses important material, makes good use of

examples and illustrations, motivates me to do my best work, inspires class confidence in his knowledge of the subject, and gives me new viewpoints or appreciations.

Using a slightly different approach, Bousfield (21) devised a checklist of 19 qualities and students rated each on a 0 to 10 basis (no consequence to highest importance) in terms of its importance to the college teacher. Although some differences existed between men and women with regard to the qualities rated most important, results indicated that most students were primarily concerned with being treated fairly and with the pedagogical competence of their professors. In a similar study by Quick and Wolfe (47), students were given a list of ten items and asked to select the three qualities they considered most important in the ideal teacher. Results indicated that the ideal professor should: encourage independent thinking, have a deep and sustained enthusiasm for his subject, and have subject matter and course well organized.

Hoffmann (32) studied the responses of college seniors to an evaluation instrument designed to help the administration of the college select a recipient for the annual outstanding-teacher award. The data showed that students valued as most important such characteristics as sincerity, patience, understanding, interest, sympathy, respect, trust and fairness. In a similar investigation, Williams (59) found that cultural interest, enthusiasm, friendliness, and slowness in speech and movement were traits which could be used to identify good teachers.

deBriun (25) conducted an investigation in which 268 graduate students rated teachers on overall teaching ability and aspects of their self-concepts. Results from this study indicated that necessary prerequisites to effective teaching were: good perception of the subject matter, sensitivity to the needs of the students, the belief that students have the ability to comprehend the subject matter, trust and respect for students as individuals, enthusiasm, and self-confidence about the subject. The data from several studies summarized by Finn (26) supported the results of the deBriun investigation.

Musella and Rusch (45) studied responses of 394 college seniors who were given a list of 10 teacher qualities and asked to identify the three most important to good teaching in the physical sciences, three in the social sciences, and three in the arts. The five qualities generally identified as being most important to effective teaching were: expert knowledge of the subject matter, systematic organization of subject matter, ability to explain clearly, enthusiastic attitude toward the subject, and ability to encourage thought. However, the data from this study indicated that "characteristics associated with effective teaching were found to be different for the physical and biological sciences than for the arts and the social sciences" (45:140). In a study conducted by Ryans (10), data collected concerning traits of teachers who ranked high and low on scales of desirable teaching characteristics indicated, too, that the pattern of traits was not the same for teachers of different subject matter.

Characteristics generally mentioned, however, as distinguishing the good teacher were: emotional stability, friendliness, cooperativeness, agreeableness, restraint, and objectivity.

A study by Morton (44) revealed that perceptions of college students regarding desirable characteristics of teachers varied depending upon the age and experience of the students, class rank, and gender. While groups of freshman men and women both agreed that knowledge of subject, clarity, and helpfulness were important qualities of good teachers, men emphasized the importance of fairness and well defined goals, while women were more concerned with personal appearance. A similar study of seniors indicated that they associated creativity and interesting style with good teaching. Senior men further stated that the comprehensiveness in competence of the teacher and the relevance of material highly influenced the quality of teaching, while senior women noted concepts such as understanding, skill in outlining and reaching high goals, and helpfulness as important teaching qualities.

A study by Harristhal (63), specifically related to physical education, revealed results similar to those of the investigations previously reported. She used a student reaction inventory to survey opinions of students regarding the competencies of women physical educators in the basic program. Results indicated that the following factors were related to teaching effectiveness: knowledge of the subject, interest in individual

students, impartiality, patience, friendly but firm leadership on an adult basis, enthusiasm for the subject, and skill in the activity taught.

Summary. A review of the results of over 30,000 opinions of pupils revealed that there is a general consensus among students regarding the characteristics of effective teachers. Concepts listed most frequently as characterizing effective teachers were enthusiasm, fairness, knowledge of subject, clarity, friendliness, interest in students, and sense of humor.

Perceptions of Students and Teachers

Several investigations have been made in which students and teachers have been asked to identify characteristics of effective teachers. Krupka (67) conducted a study in which faculty and students rated 12 areas of an Instructor Rating Questionnaire in terms of their importance in judging a teacher. Both groups ranked the 12 areas from 1 to 12, most important to least important. Each group ranked knowledge of the subject and ability to arouse interest in students as first and second in importance, and both faculty members and students ranked organization of the course, classroom presentation, and the teacher's willingness to help in the six most important categories. These results indicated that there existed a high positive relationship between the way faculty and students judged the areas utilized in the study.

Isaacson, McKeachie, and Milholland (33) correlated peer ratings of 33 teaching fellows with ratings by students and found that those teachers rated by their peers as being artistically

sensitive, emotionally stable, intellectual, poised, energetic, and imaginative tended to be the ones rated as effective instructors by their students. Data from a similar study by Maslow and Zimmerman (39) showed that students tended to equate good teaching with good personality, while colleagues tended to equate good teaching with creativeness. The findings revealed, however, that students and faculty agreed fairly well on who the good teachers were ($r = .69$).

Using a slightly different approach, Hildebrand and Wilson (65) administered several surveys in which 338 students described their best and worst teachers and 119 faculty members described the teaching of colleagues whom they regarded as best and worst. Results revealed that excellent agreement existed among students, and between faculty and students about the effectiveness of given teachers, and that items listed as characterizing best teachers as perceived by students and by colleagues did discriminate statistically the best from the worst teachers at a high level of confidence. In a similar study, Yourglich (61) found that, although students and faculty did not agree consistently ($r = .59$) on the characteristics of the ideal teacher, the four qualities ranked highest by both groups were integrity, understanding, ability to communicate, and maturity.

Cole (4), summarizing 23 studies related to opinions of students, faculty members, administrative officers and alumni regarding traits of good and poor teachers, reported results similar to those of the studies previously mentioned. He noted

the following concepts as characterizing the outstandingly good teacher: knowledge of the subject, organization, preparation, clarity, fairness, informal class atmosphere, friendliness, enthusiasm, and sense of humor.

Summary. It appears from the literature that there is general agreement between faculty members and students with respect to essential characteristics of effective teaching, and regarding the effectiveness of given teachers. Concepts mentioned in studies involving opinions of both teachers and pupils included knowledge of subject, organization, enthusiasm, fairness, and friendliness.

Perceptions of Teachers

Chiu (23) conducted a study to determine the characteristics of effective teaching solely as perceived by teachers. One hundred eighty-four subjects were asked to describe the essential characteristics of effective teaching. Items listed most frequently were: individualization of instruction, ability to motivate students to learn, good discipline, involvement of pupils in the class, adequate preparation, knowledge of individual pupils, interest in children, enthusiasm about subjects taught and teaching in general, a free and relaxed atmosphere, rapport with students, and adequate knowledge of subject.

Summary

A review of the literature reveals that there is general agreement among students and between students and teachers regarding qualities necessary to effective teaching, and with respect to

the effectivensss of given teachers. While the importance of certain characteristics may vary according to the subject matter being taught, and/or according to the nature of the student, concepts reported consistently in the research as being critical to the teaching process were: clarity, enthusiasm, fairness, friendliness, interest in students, knowledge of subject, and sense of humor.

TEACHER BEHAVIOR

Teacher Behavior and Student Achievement

A review of the research indicated that few studies have been conducted to determine the relationship between characteristics or behaviors of teachers, and student achievement (11, 43, 50).

In a study by Ryans (51), attention was focused on a limited number of important teacher behavior dimensions in order to determine their relationship to student achievement. Bipolar dimensions of teacher behavior judged to be important in the classroom included: partial-fair, autocratic-democratic, aloof-responsive, restricted-understanding, harsh-kindly, dull-stimulating, stereotyped-original, apathetic-alert, unimpressive-attractive, inarticulate-articulate, monotonous-pleasant, evading-responsible, erratic-steady, excitable-poised, uncertain-confident, disorganized-systematic, inflexible-adaptable, pessimistic-optimistic, immature-integrated, and narrow-broad. Trained observers employed these dimensions by assigning each teacher observed a value, on a scale from one (at the left pole) to seven (at the right pole) on each dimension. For secondary

school classes, only the teacher behavior dimension which described the extent to which a teacher was judged dull-stimulating seemed to be closely associated with productive pupil behavior.

Assuming that the concept "dull-stimulating" related somewhat to the enthusiasm of the teacher, the finding of Ryans' study was supported by data from two other studies on teacher behavior and student achievement. Smith (11), in a discussion of the results of the process-product studies conducted prior to 1971, reported that clarity, variability, businesslike behavior, and student opportunity to learn, in addition to enthusiasm, showed the strongest relationship to student achievement. In a similar paper summarizing the results of high-inference studies, Rosenshine (49) reported that ratings given to teachers on "such behaviors as stimulating, energetic, mobile, enthusiastic, and animated were related to measures of student achievement" (49:510). In addition, he noted that the frequencies of such specific behaviors as eye contact, variation in voice, gesture, and movement were related to student achievement.

Summary. The results from the above studies relating various teacher behaviors to measures of student achievement indicated that, of the concepts recognized earlier in this chapter by students and teachers as reflecting teacher effectiveness, enthusiasm was most crucial to productive pupil behavior, while clarity also proved important. A summary of the concepts noted in the aforementioned studies appears in Table 1.

Table 1

Summary of Concepts Mentioned in Studies Regarding Characteristics of Effective Teachers

	Clarity	Enthusiasm	Fairness	Friendliness	Interest in Students	Knowledge of Subject	Sense of Humor
Barr		X					X
Bousfield			X			X	
Chiu		X			X	X	
Cole	X	X	X	X		X	X
Davis	X		X	X		X	X
deBriun		X					
Finn		X			X	X	
Harristhal		X	X	X	X	X	
Hart	X	X	X	X	X	X	
Hoffmann			X		X		
Isaacson, et al.		X				X	
Krupka						X	
Langen	X						
MacDonald			X	X	X	X	X
McComas		X	X		X	X	X
Morton	X		X			X	
Musella and Rusch	X	X				X	
Quick and Wolfe		X					
Rosenshine		X					
Ryans		X	X	X			
Schaffle	X		X		X	X	
Smith	X	X					
Tiedeman	X		X	X			X
Williams		X		X			
Yourglich	X						
N = 25	10	15	12	8	8	14	6

Studies of Verbal Behavior

Several efforts have been made to construct descriptive scales based on verbal behavior in the classroom.

Flanders (6), assuming that verbal behavior was an adequate sample of the teacher's total behavior pattern, used information from research data regarding the identification of different kinds of verbal statements made by teachers to develop a 10-category system for describing classroom interaction. Seven of the categories were assigned to teacher talk, two were relegated to student talk, and one covered short periods of silence, noise, or confusion. Of the seven categories assigned to teacher talk, four represented actions which increased the active control of the teacher, and three described actions which afforded greater freedom for students. The four categories describing teacher talk, indirect influence, were: accepts feeling, praises or encourages, accepts or uses ideas of students, and asks questions. Teacher talk, direct influence, categories included: lectures, gives directions, and criticizes or justifies authority. Student talk categories were labeled student talk-response and student talk-initiation. Each of the categories was numbered and was described in greater detail on the scale.

This system was used by highly trained observers who sat in the classroom and, at the end of each three-second period, wrote down the category number which best represented the communication event just completed. Those numbers were then plotted on a 10 X 10

matrix and, combined with a general description of the activity period, provided a graphic view of the teacher's total pattern of influence.

Bellack and Davitz (8) were concerned, too, with the verbal interaction which characterizes classrooms in action. They studied tape recordings of Problems of Democracy classes and decided that the verbal interplay between teachers and students fell into four types of pedagogical moves. These categories were labeled: structuring, soliciting, responding, and reacting. Structuring moves included verbal statements which focused attention on the topic of the day or set the context for subsequent behavior. Soliciting moves encompassed statements or questions designed to elicit responses, while those statements made in response to soliciting moves were classified in the category labeled responding. Moves designated as reacting included those verbal statements intended to modify or evaluate what had been said previously. While examining transcripts of classroom discussions, pairs of coders identified each type of move as it occurred in the discourse and coded it according to the appropriate category number.

In a similar effort by Withall (57), seven categories were identified as describing teacher verbal behavior in the classroom. Three learner-centered categories, called learner-supportive, acceptant and clarifying, and problem-structuring statements, and three teacher-centered categories, labeled directive or hortative, reproving or deprecating, and teacher self-supporting remarks comprised six of the seven categories developed. The final category

was designed for neutral statements. This instrument was used by applying the seven categories to teacher statements to determine whether the pattern of verbal behavior was primarily learner-centered, teacher-centered, or problem-centered. Although no statistical evidence was given, the scale was reported to have validity, objectivity, and reliability.

A system developed by Smith and Meux (70) was designed to describe classroom verbal behavior by determining the nature of the opening phase of each verbal move. A new verbal move occurred each time there was a shift in what the speakers were talking about. This scale was constructed with the assumption that initial verbal moves, called entries, tended to shape the character of episodes occurring in a classroom discourse. The categories into which the entries were grouped were: (1) defining, (2) describing, (3) designating, (4) stating, (5) reporting, (6) substituting, (7) evaluating, (8) opining, (9) classifying, (10) comparing and contrasting, (11) conditional inferring, (12) explaining, and (13) directing and managing the classroom. Although data for classes in English, science, mathematics, and social studies indicated that the most frequently occurring operations were describing, designating, and explaining, and the least frequent were substituting, reporting and classifying, the writers concluded that the extent to which the various operations are employed probably varies from teacher to teacher, and from subject area to subject area.

Ryans (52), describing the teacher as an "information-processing system," assumed five major categories into which

teacher behaviors fell: motivating-reinforcing, presenting-explaining-demonstrating, organizing-planning-managing, evaluating, and counseling-advising. Specific information was not reported regarding the development of the categories or use of the system in the observation and description of teacher behavior.

Studies of Nonverbal Behavior

Lewis, Newell, and Withall (36) attempted to utilize both verbal and nonverbal behaviors in the classroom in order to study the process of communication between teachers and students. They constructed a scale of 13 verbal and nonverbal categories based on the inferred intent of the communicator. These categories were entitled: (1) asks for information, (2) seeks or accepts direction, (3) asks opinion or analysis, (4) listens, (5) gives information, (6) gives suggestions, (7) gives direction, (8) gives opinion, (9) gives analysis, (10) shows positive feeling, (11) inhibits communication, (12) shows negative feeling, and (13) no communication. Observers using this scale assigned one or more category numbers to each 10-second time interval, and summarized the total communication pattern by tabulating the frequencies for each category.

In an attempt to focus exclusively on the nonverbal behavior of teachers, Galloway (29) designed seven categories which could be utilized to describe the inferred intent of each communicative act which occurred in the classroom. The three categories which were designated as describing encouraging

communications were labeled enthusiastic support, helping, and receptivity. Categories entitled inattentive, unresponsive, and disapproval were designed to encompass inhibiting communications. The final category, pro forma, was considered neither encouraging nor inhibiting. When a communicative act occurred which was related to this category system, the observer recorded a number representative of the appropriate category. In addition to this, observers wrote descriptions of the physical setting of the classroom, the communicative acts which occurred, the contextual situation of the teacher's behavior, and the "how" of everything done by both teacher and students. The categorization of communicative acts, in combination with general descriptions of the classroom, provided an overall view of classroom climate.

Summary. A review of the literature revealed that numerous attempts have been made to devise scales which describe teacher behavior and classroom climate. Although nonverbal behavior was considered in two of the studies, most of the scales were categorical systems designed to describe the degree to which the verbal behavior of the classroom was teacher-centered or student-centered. The majority of systems employed a method of coding communicative acts at set intervals in the discourse, in addition to a general description of the climate of the classroom. All of the systems required a high level of training on the part of the raters using them.

SCALE CONSTRUCTION

Most rating scales which have been constructed on the basis of characteristics of good and poor teaching have taken one of three forms: (1) a list of questions that an observer answers positively or negatively, (2) a list of traits for each of which an observer rates the degree to which the teacher shows said trait, and (3) a list of activities which an observer uses to indicate what the teacher actually does and how often he does each (4). Instruments which have taken one of the first two forms have been primarily oriented toward evaluating teacher behavior, while scales which have been based on a list of activities and their frequency of occurrence have been, basically, descriptive instruments. Since the purpose of the present study was to identify, not value, teaching behaviors, this report on scale construction has been limited to scales which are designed to describe teacher behavior and/or classroom climate, and to one study which reported procedures similar to those used in the present study.

Teacher Characteristics and Behavior

In an early study, Beecher (1) attempted to develop a teacher rating scale based on observable teacher behavior items characteristic of what students said they liked in teachers. The following steps were taken in the development of the scale: (1) analysis of studies in which students reacted to behaviors of teachers, (2) formulation of a scale utilizing the behaviors listed most frequently, (3) preliminary experimentation in the use and

refinement of the scale, (4) establishment of reliability of the experimenter's ratings using the scale, (5) establishment of validity of the scale in terms of criterion judgments of experts, and (6) establishment of reliability of the scale using two judges rating the same individuals.

On the basis of studies involving over 30,000 pupil reactions to desirable teaching qualities and behavior, Beecher selected the following categories for a teacher rating scale: (1) indications of fairness, (2) indications of cheerfulness, (3) indications of sympathetic understanding, (4) indications of control, (5) indications of ability to get pupil response, and (6) indications of knowledge and skill. The investigator then selected teacher behaviors most frequently associated by students with the above characteristics and grouped them with the appropriate categories. The following are selected examples: (1) indications of fairness: praise and criticism based on fact, no favoritism shown, no excessive criticism of individual pupils; (2) indications of cheerfulness: happy facial expression and voice, never shows impatience, shows sense of humor, friendly in manner and tone to all pupils; (3) indications of control: objectives clear to teacher and pupil, insistence on order and obedience, evidence of thorough planning, authority unquestioned; (4) indications of ability to get pupil response: personal enthusiasm, well-inflected voice, inspires eager responses from pupils, encourages response and individual comment; and (5) indications of knowledge and skill: explains lessons and answers

questions clearly and thoroughly, creates interest through example, and provides for repetition.

The scale was then submitted to a group of experienced state supervisors, superintendents, principals, teachers, and pupils for suggested changes. After preliminary application of the scale, both in its original form and with various suggested changes, the items first selected were retained.

Further validation of the instrument was accomplished by correlating results from use of the instrument with ratings of the same teachers by members of a panel of experts comprised of state supervisors, supervising principals, and superintendents of schools. The criterion judges rated each of 50 teachers as superior, average, or poor, and submitted the ratings to the investigator after he had applied the scale at least twice to each subject. The resulting coefficient of .88 indicated a significant degree of validity for the instrument. A check of the reliability of the instrument revealed a correlation coefficient of .79, indicating significant consistency between judges independently rating the same teachers.

Analyses of Classroom Climate

Many of the scales which have been constructed regarding the teaching process have been based on the assumption that verbal communication constitutes an adequate sample of classroom climate and of the teacher's total pattern of influence. Most investigators have focused on developing a set of categories after studying the nature of classroom interaction, determining a proper unit of

analysis, and establishing procedures for the use of the category system developed.

Flanders (6) developed a system of 10 categories designed to describe teacher talk, pupil talk, and the degree of teacher control or student freedom by classifying, at three-second intervals, verbal statements in the classroom. Teacher talk categories included: (1) accepts feeling, (2) praises or encourages, (3) accepts or uses ideas of student, (4) asks questions, (5) lectures, (6) gives directions, and (7) criticizes or justifies authority. Student-talk categories were labeled student talk-response, and student talk-initiation. The final category was used for periods of silence or confusion. All of the categories were further defined for use by the observers. This system of interaction analysis also included identification by the observers of various activity periods in the classroom. At the completion of each observation, the rater, in addition to summarizing the categorization of verbal statements, wrote a general description of each activity period.

Bellack and Davitz (8) studied transcripts of discussions of Problems of Democracy classes and designed four major categories into which they suggested the verbal interplay of students and teachers could be classified: (1) structuring, (2) soliciting, (3) responding, and (4) reacting. Each of these categories was described in detail and coders marked each change in the classroom discourse according to the category most descriptive of it.

Computations for agreement between coders ranged from 84 to 94 per cent indicating that the system of analyzing classroom discourse was objective for Problems of Democracy classes.

In a study similar to the Flanders and Bellack and Davitz studies, Withall (60) decided that teacher-statements tended to fall into about 25 types of responses. Since they were found to overlap, the categories were reduced to seven which seemed to encompass all of the kinds of statements which teachers utilized in the classroom. These categories were labeled: (1) learner-supportive, (2) acceptant and clarifying, (3) problem-structuring, (4) neutral, (5) directive or hortative, (6) reproving or deprecating, and (7) teacher self-supporting remarks. This instrument was employed by coding each teacher-statement according to the category which best described it. The mean percentage of agreement for judges applying the scale to several typescripts was 65 per cent.

Smith and Meux (70) designed two methods of describing classroom discourse. First, they determined that an entire discourse could be divided into two kinds of units: episode and monolog. Subsequently, they developed a classificatory scheme for episodes based on the verbal move contained in each opening phase of an episode. A change in episode was determined not by a shift in speakers, but by an alteration in what the speakers were talking about. Each new verbal move, called an entry, was placed in one of the following categories: (1) defining, (2) describing, (3) designating, (4) stating, (5) reporting,

(6) substituting, (7) evaluating, (8) opining, (9) classifying, (10) comparing and contrasting, (11) conditional inferring, (12) explaining, and (13) directing and managing the classroom. The median correlation coefficient for agreement between pairs of judges utilizing episode and monolog as the units of analysis was .70. The median coefficient for objectivity using the system of 13 categories was .67.

Lewis, Newell, and Withall (36) used the Bales Interaction Process Analysis categories as a core and, after redefinition, deletion, and addition of classifications, constructed 13 verbal and nonverbal categories which could be applied to the process of communication in the classroom. The categories, based on the inferred intent of the communicator, were (1) asks for information, (2) seeks or accepts direction, (3) asks opinion or analysis, (4) listens, (5) gives information, (6) gives suggestions, (7) gives direction, (8) gives opinion, (9) gives analysis, (10) shows negative feeling, (11) inhibits communication, (12) shows positive feeling, and (13) no communication. A 10-second time interval was used as the unit of analysis, and the observer assigned a category score, or more if necessary, to each time unit. At the completion of an observation, a communication pattern was summarized by tabulating the frequencies for each category. No coefficient for objectivity was reported.

Using a different approach, Ryans (51) grouped specific individual behaviors into general classifications in order to study the nature of the student behavior-teacher behavior

relationship. For secondary school classes, he identified 21 bipolar dimensions of teacher behavior, such as autocratic-democratic, dull-stimulating, and disorganized-systematic. Trained observers employed the dimensions by applying a rating of one to seven (left pole to right pole) to each teacher. The assessment procedure was somewhat standardized by the use of a glossary which accompanied the assessment form and was utilized in training judges. Correlation coefficients computed for objectivity clustered around .60.

In one of the few attempts to deal exclusively with the nonverbal dimension of teacher behavior, Galloway (29), after conducting a study of observation procedures for determining teacher nonverbal communication, constructed a seven-category scale to enable observers to make inferences from the nonverbal behavior of teachers. Three categories, labeled enthusiastic support, helping, and receptivity, were considered to be encouraging communications; categories labeled inattentive, unresponsive, and disapproval were considered inhibiting, and the final category, pro forma, was considered neutral. All categories were described in detail. With this system, when a communicative act occurred which was related to the category system, observers recorded a number representative of the category. Objectivity for this system was not reported.

SUMMARY

Various studies have been conducted to determine characteristics of effective teachers. Results have indicated a number

of broad concepts which appear to be related to teaching effectiveness. It appears from the research, however, that these characteristics, exemplified by such terms as enthusiasm, clarity, fairness, interest in students, knowledge of subject matter, friendliness, and sense of humor lack operational definitions.

Various descriptive scales of verbal and nonverbal classroom behavior have been developed from analyses of communicative acts by teachers and students. Most of these category systems focus on teacher verbal behavior, and are designed to describe the overall climate of the classroom.

Regarding teacher and student verbal and nonverbal behavior, most scale construction has been done by first, finding a basis for and developing a system of categorization; second, deciding upon a unit of analysis; and finally, determining overall procedures and means of summarizing the observations.

CHAPTER III

PROCEDURES

The procedures used by the investigator in the present study included the following: (1) design of the study, (2) development of a trial scale, (3) pilot study, (4) revision of the trial scale, (5) collection of data, and (6) analysis of data.

DESIGN OF THE STUDY

The present study was designed to develop a scale which may be used to observe and describe the frequency of occurrence of selected teacher behaviors in physical education activity classes. The behaviors on the final scale were chosen after an analysis of the literature, introspection, and validation by a jury of experts. The operational validity of the items was also checked by interviewing three observers after they had utilized the scale to describe the behavior of five different instructors teaching various college physical education activity classes.

DEVELOPMENT OF TRIAL SCALE

As characteristics of effective teaching were identified through an analysis of the literature, they were recorded on cards and tallied on a chart. The reader is referred to Table 1,

page 16, in Chapter II. It was decided that, of the concepts appearing in the research, not less than five nor more than ten would be selected for use in the present study. Therefore, only those qualities appearing repeatedly in the studies reported were charted. A summary of the research revealed seven concepts which were listed frequently as reflecting effective teaching. They were: (1) enthusiasm, (2) knowledge of subject, (3) fairness, (4) clarity, (5) friendliness, (6) interest in students, and (7) sense of humor. These seven concepts were chosen for the purposes of the present study.

Having identified the broad concepts to be used in the trial scale, the investigator then employed further documentary analysis and introspection in order to determine specific behavioral correlates of each of these concepts. During the entire course of the preliminary investigation, a list of possible behavioral items was in constant development. The primary purpose was to develop a list of specific, observable, behavioral items which were valid correlates of the concepts selected for use in the study. During this process, items were combined, restated, or deleted in order to achieve comprehensiveness, while avoiding repetition or undue length. Thirty-one teacher behaviors emerged at the conclusion of this procedure and were classified arbitrarily by the investigator under the seven broad concepts.

Since the purpose of the scale was to describe, not value, the frequency of occurrence of the behaviors listed, the following three categories were selected as the units of analysis:

(1) frequently or always, (2) sometimes, and (3) seldom or never. The number of categories was limited to three in order to minimize ambiguity. Because of the relative nature of many of the items on the scale to the teaching process, no attempt was made by the investigator to define specifically the terms frequently or always, sometimes, and seldom or never. A copy of the trial scale appears in Appendix C.

PILOT STUDY

The procedure for the pilot study is reported in three sections: (1) jury of experts, (2) observations of teachers, and (3) revision of trial scale.

Jury of Experts

A jury of experts was selected by the investigator to aid in the validation and revision of the trial instrument.

Selection of jury members. Ten jury members were selected on the basis of the following minimum requirements: (1) Ph.D. or equivalent, (2) 10 years or more of teaching experience at the college level, and (3) present rank of Associate Professor or Professor. These criteria were employed since they are generally recognized as representative of expertise in a given area.

The jury members were faculty and administrative personnel at five different universities who were identified as possessing extensive training and knowledge in the areas of scale construction, teacher behavior, supervision of teachers in physical

education, the behavioral bases of physical education, and administration of physical education. Since the purpose of the study was to develop a scale for observing and describing the frequency of occurrence of selected teacher behaviors, it was decided that experts in the areas of teacher behavior and scale construction might be able to make critical comments and valuable suggestions about the initial instrument. Experts in the area of supervision of teachers were asked to assist because of their practice and skill in observing teacher behavior, and persons with extensive background in the behavioral bases of physical education were included because of their ability to examine the items as related specifically to the area of physical education. The assistance of administrators was solicited because of their concern regarding teacher characteristics and behavior as they relate to selection and retention procedures, improvement of teacher preparation programs, and administrative decisions regarding academic rank, tenure, salary, and merit raises.

Although all ten persons selected for the jury agreed to participate in the study, only nine returned the checklists. A list of the jury members appears in Appendix A.

Contribution of jurors. The members of the jury were asked to react to the trial instrument by indicating whether or not they thought the items selected for the scale were: (1) behavioral correlates of any of the concepts listed, and (2) observable. Prior to sending the checklist to jury members, the behavioral items to be classified were randomized utilizing

a table of random numbers, and the concepts into which the behaviors were to be categorized were placed in alphabetical order (12). This procedure was followed in order to obviate any patterning in responses. Copies of the directions to jury members and the scale used by jurors are included in Appendix A.

Results. Prior to the beginning of the pilot study, it was decided that, in order for an item to be retained on the final scale, it was necessary for at least six of the jury members to place it under one of the seven concepts on the scale and indicate that it was an observable behavior. A summary of how jury members classified the items appears in Appendix A.

Of the thirty-one items on the trial scale, four were eliminated by jury members: (1) allows students to tell jokes or play practical jokes, (2) changes voice inflection, (3) dresses appropriately for the activity, and (4) uses gestures. The remaining 27 items were classified by at least six experts and were judged to be observable. Sixteen additional items suggested by jurors were reviewed by the investigator and the observers who had been practicing with the scale to determine whether or not the items were clear, unrepetitive of items already on the scale, and practical in terms of operational use. Three were selected for inclusion in the final study: (1) maintains good eye contact with students, (2) laughs at self when appropriate, and (3) gets students actively involved in learning early in the lesson.

The general comment occurring most frequently on the returned forms related to the difficulty jurors had in placing many of the items under only one concept. For this reason, some jury members failed to classify all of the items or, on occasion, indicated two concepts with no first and second choices. In those cases, the item was not tallied for that juror. If choices were given, the item was tallied in the first choice category.

Because jury members had difficulty discriminating between some of the categories, and the observers indicated that the organization of category headings was not crucial to them, the decision was made to combine six of the seven category headings into pairs, and group the tallies under those concepts together in order to determine the classification of items for the final scale. Thus, in order to eliminate ambiguity and overlap, the cognitive qualities of clarity and knowledge of subject were placed in one category. Friendliness and interest in students, qualities which denote association with students, were united, and the personal characteristics of enthusiasm and sense of humor were combined. The final single category was fairness. For the purposes of the final study, each behavioral item was placed in the category into which the majority of jurors said it belonged.

Observations of Teachers

At the same time the jury of experts was validating the trial scale, a study was being conducted to determine the

objectivity of rating the frequency of occurrence of the items on the scale. Three raters observed five different teachers during three consecutive instructional class sessions and marked the scale independently. Ratings were made separately for each day, and a composite rating was completed for each teacher at the end of the third day. This composite form was filled out without reference to the first three ratings. The decision to utilize a three-day procedure was made in order to obtain more data and to study the objectivity of a composite rating. For the purposes of the observations, the concepts were placed on the forms in alphabetical order and the behavioral items were classified arbitrarily by the investigator. Each item was judged as occurring frequently or always, sometimes, or seldom or never. A copy of the rating form utilized by the observers appears in Appendix C.

The teachers observed during the pilot study were instructing college classes in beginning swimming, advanced swimming and lifesaving, body mechanics, gymnastics, and modern dance. Several different activities were included in order to study the objectivity of judging the frequency of teacher behaviors under different circumstances. The classes were not selected randomly but were chosen on the basis of accessibility.

Prior to the start of the observations, individual conferences were held with the teachers involved in the study to explain the purpose of the observations and to answer any questions. In addition to the conferences, a written reminder was given to each teacher informing him of when the observations would begin.

During both of these procedures, emphasis was placed on the importance of the classes being instructional in nature. Teachers were asked to suggest three consecutive class sessions during which no examinations or films would be scheduled and during which instruction would occur. The schedule of observations was established accordingly. A copy of the memorandum appears in Appendix B. The classes met for 50 minutes, three times per week. All classes were observed within a two-week period of time.

Results from the pilot study were analyzed by using the Pearson product-moment correlation method, and Fisher's *z* method for averaging correlation coefficients (7). Coefficients for 60 ratings computed between two independent observers ranged from .43 to .93, with an overall coefficient of .73. The highest daily coefficient (.79) existed for Day 3, while the coefficient for Day 1 was lowest (.66). The reader is referred to Table 2 for a summary of the results of the objectivity study. The resultant coefficient of correlation for all the observations, .73, indicated that fairly high agreement existed between observers regarding the frequency of occurrence of the thirty-one behaviors on the trial scale (13). The agreement between Judges 1 and 2 was .77, while resulting coefficients between Judges 1 and 3, and 2 and 3, were .73 and .69 respectively.

The investigator met with the observers at the conclusion of the pilot study to discuss the operational validity of the items on the scale. They suggested that five of the items be separated in order to avoid ambiguity. These were items numbered

Table 2

Objectivity Coefficients for Rating the Frequency of Occurrence of 31 Behavioral Items

Day	Observers	T_1[a]	T_2[a]	T_3[a]	T_4[a]	T_5[a]	Overall[b]
	1 - 2	.60	.57	.55	.86	.68	.67
1	1 - 3	.56	.71	.51	.77	.71	.66
	2 - 3	.57	.58	.57	.71	.78	.65
	1 - 2	.43	.79	.79	.67	.78	.71
2	1 - 3	.61	.72	.72	.80	.86	.75
	2 - 3	.76	.68	.77	.71	.62	.71
	1 - 2	.80	.77	.89	.93	.84	.86
3	1 - 3	.76	.84	.73	.72	.65	.75
	2 - 3	.84	.76	.72	.69	.67	.74
	1 - 2	.74	.84	.81	.85	.78	.81
Comp.	1 - 3	.69	.92	.66	.71	.56	.74
	2 - 3	.72	.80	.64	.58	.55	.67
Overall[b]:		.69	.76	.71	.77	.72	.73

[a]Computed by using the Pearson product-moment method of correlation.

[b]Computed by using Fisher's *z* method of averaging correlation coefficients.

Judges 1 - 2 = .77
Judges 1 - 3 = .73
Judges 2 - 3 = .69
Day 1 = .66
Day 2 = .72
Day 3 = .79
Comp. = .74

T_1 = Modern Dance
T_2 = Beginning Swimming
T_3 = Body Mechanics
T_4 = Advanced Swimming and Lifesaving
T_5 = Gymnastics

8, 14, 24, 27, and 29. (See Appendix C for content of items.) They also suggested that, since the purpose of the scale was descriptive rather than evaluative, items 13 (avoids excessive criticism of students) and 14 (gives individual attention to all students and avoids spending an excessive amount of time with one or a few) be stated in positive rather than negative terms. A further suggestion was to eliminate the words "all" from item 14, and "each" from item 21 (verbally encourages each student individually).

All observers indicated that it was easier to use the scale in the swimming and gymnastics classes than in body mechanics or modern dance sessions. This may have been due to the fact that the former classes are usually more structured than the latter. Although the investigator can only speculate as to the differences in objectivity for these classes, it is interesting to note that the comments of the observers were supported by the sizes of the resultant correlation coefficients for the classes observed. It was the opinion of the raters, at the conclusion of the pilot study, that the scale would be more easily used in team sports classes than in individual or dual activity courses.

The observers indicated that three observations seemed to be adequate prior to filling out the composite form. They noted, however, that concentrating all observations in such a short period of time may have impaired their ability to make clear judgments on the composite ratings.

At the conclusion of the pilot study, the investigator also discussed with the observers each of the behaviors which had been suggested by the jury members. In terms of practicality of use on the scale, only three behaviors were approved unanimously by the observers: (1) maintains good eye contact with students, (2) laughs at self when appropriate, and (3) gets students actively involved in learning early in the lesson.

Revision of Trial Scale

After receiving the completed forms of the jury members, studying the data from the study of objectivity, and reviewing the suggestions and comments of the observers, the following revisions were made prior to the final study:

1. Five behaviors were divided into separate items.
2. Four behaviors were eliminated from the scale.
3. The wording of two behaviors was modified slightly.
4. Two behaviors were changed from negative to positive statements.
5. Six of the seven concepts were placed in combination form, and behaviors for the combined concepts were grouped together.
6. Three new items were added to the scale. They were classified arbitrarily by the investigator under one of the four broad categories.

The final form contained 35 items. A copy of the revised scale appears in Appendix C.

COLLECTION OF DATA

Procedures

During the final study, three observers rated five teachers during three consecutive instructional class sessions on each of the 35 items on the scale. Because the observers were the same for the final study as for the pilot study, they had practiced using a version of the form and were generally familiar with it.

For each class session the observers completed a rating form independent of the other raters. Each item was marked as occurring frequently or always, sometimes, or seldom or never. In addition, a composite form was completed at the end of the third day. This rating was made without reference to the other forms. A total of 60 ratings between paired observers was obtained.

The college teachers observed were instructing classes in archery, body mechanics, beginning swimming, tennis, and track and field. The classes were not chosen randomly but were selected on the basis of accessibility. The investigator met with each teacher in advance in order to clarify the purpose of the observations and to answer questions. In addition, a written memorandum was distributed to the teachers to inform them of the scheduled times during which observations would be made. During the preliminary conferences, emphasis was placed on the importance of the classes being instructional in nature. Teachers were asked to suggest three consecutive periods during which examinations and

films would not be scheduled and during which some teaching would occur. The schedule for the observations was established accordingly. Copies of the memorandum to the teachers, and the schedule for observations, are included in Appendix B.

Four of the five classes met three times a week for 50 minute periods; the track and field class met twice a week for 75 minutes each session. The latter class was included to study the possible effect of a longer observational period on the objectivity of the scale. All of the classes were observed within a three-week period of time.

Data Collected

Following the three-week observational period, the following data were collected:

1. Twenty rating forms from each observer, four for each of the five teachers.
2. Information from the observers regarding the operational validity of the revised scale.
3. Suggestions from the observers with respect to the general use of the scale.

ANALYSIS AND DISCUSSION OF DATA

The objectivity of rating the frequency of occurrence of the items on the scale was statistically analyzed by using the Pearson product-moment correlation method, Fisher's *z* method for averaging correlation coefficients, and the percentage of agreement method.

The first two statistical techniques were utilized to determine objectivity between observers for the total scale. The percentage of agreement method was employed to determine objectivity on each of the 35 items. It was computed by: (1) assigning values of 3, 2, and 1 to the units of analysis (frequently or always, sometimes, and seldom or never); (2) tallying the degree of disagreement for each observation by marking a 2 if one judge said frequently or always while another said seldom or never, and a 1 if paired observers marked any two adjacent categories (no tally was made for agreement); and (3) subtracting the number of tallies indicating disagreement from the total number of paired observations (N = 20 between two observers, N = 60 overall) and dividing by N to determine the percentage of agreement. For example, if the total of the tallies for disagreement on an item between all judges was 14, the percentage of agreement was computed by subtracting 14 from 60, and dividing the resultant number, 46, by 60, to report a percentage of agreement of 77 per cent.

In addition to the statistical analysis, information obtained in a final conference with the observers was reported and discussed in detail.

SUMMARY

The present study was designed to develop a scale which may be used to observe and describe the frequency of occurrence of selected teacher behaviors in physical education activity classes.

In order to develop a trial scale, concepts reflecting effective teaching were identified through an analysis of the literature, and behavioral correlates of these concepts were developed by documentary analysis and introspection. A total of seven concepts and 31 behavioral correlates resulted. During a pilot project, the objectivity and validity of the trial scale were studied.

Validity was established by submitting the trial scale to a jury of nine experts in physical education. Jurors marked whether or not they considered the 31 items to be observable teacher behaviors, and classified each under the concept best representing it. They also made suggestions for additional behaviors. Operational validity was studied through the use of the scale during 60 observations.

A total of 60 independent observations on five different teachers provided the data for studying the objectivity of rating the frequency of occurrence of the behaviors on the trial scale. Results revealed a correlation coefficient of .73, indicating fairly high agreement among the observers.

Suggestions of the jury members and observers were analyzed prior to revising the scale for collection of the actual data and a total of six changes resulted.

For the final study, four ratings on five different instructors were made independently by three observers. The data were analyzed by means of the Pearson product-moment method of correlation, Fisher's z method for averaging correlation

coefficients, and the percentage of agreement method. Information regarding the operational validity of the revised scale was collected, reported, and discussed.

CHAPTER IV

ANALYSIS OF DATA

The purpose of the present study was to develop a scale which may be used in observing and describing the frequency of occurrence of selected teacher behaviors in physical education activity classes. The effectiveness of the scale was determined through use, and by statistical analysis.

The statistical procedures used to analyze the data collected were the Pearson product-moment correlation method, Fisher's z method for averaging correlation coefficients, and the percentage of agreement method. These techniques were utilized to determine the objectivity of rating the frequency of occurrence of 35 behavioral items. In addition, information regarding the operational validity of the scale was collected, reported, and discussed.

PRESENTATION OF FINDINGS

Objectivity

In order to determine the objectivity of the descriptive scale developed in the present study, three raters observed each of five teachers during three instructional periods and marked the scale independently. For each day, all of the items were marked as occurring frequently or always, sometimes, or seldom or never. In addition, a composite rating for each teacher was

completed at the end of the third day of observations. This final form was marked without reference to the previous ratings.

The agreement for a total of 60 observations recorded by two independent scorers was .76. Coefficients for three judges for Day 1, Day 2, Day 3, and the composite rating were .73, .81, .78, and .73 respectively. Coefficients between Judges 1 and 2, 1 and 3, and 2 and 3 were .83, .73, and .72 respectively. Agreement between observers was highest when observing tennis and beginning swimming classes (.84 and .83) and lowest when rating teacher behaviors in an archery class (.63). A summary of the correlation coefficients computed for objectivity is reported in Table 3.

In order to determine statistically the amount of agreement between judges on each of the 35 items on the scale, the percentage of agreement was computed for each behavior. The percentage of agreement for all three observers on 35 behaviors ranged from 37 to 97 per cent. The range in percentage of agreement between two raters was 40 to 100, 25 to 100, and 20 to 95 per cent for Judges 1 and 2, 1 and 3, and 2 and 3 respectively. On all except three items, numbers 11, 17, and 25, at least two of the observers agreed 70 per cent of the time or more, and on all except four items, 9, 17, 24, and 34, the degree of agreement for all three judges was 50 per cent or greater. A summary of the percentage of agreement on each of the 35 items appears in Table 4.

Table 3

Objectivity Coefficients for Rating the Frequency of Occurrence of 35 Behavioral Items

Day	Observers	T_1[a]	T_2[a]	T_3[a]	T_4[a]	T_5[a]	Overall[b]
	1 - 2	.83	.78	.90	.90	.73	.84
1	1 - 3	.63	.64	.79	.63	.52	.65
	2 - 3	.71	.46	.80	.64	.59	.66
	1 - 2	.61	.75	.85	.93	.93	.85
2	1 - 3	.69	.72	.91	.76	.79	.79
	2 - 3	.64	.79	.91	.74	.75	.78
	1 - 2	.56	.92	.81	.93	.82	.84
3	1 - 3	.54	.76	.81	.88	.56	.74
	2 - 3	.44	.83	.79	.87	.72	.76
	1 - 2	.65	.80	.80	.82	.77	.77
Comp.	1 - 3	.64	.60	.82	.83	.73	.74
	2 - 3	.45	.64	.70	.83	.62	.67
Overall[b]:		.63	.75	.83	.84	.73	.76

[a]Computed by using the Pearson product-moment method of correlation.

[b]Computed by using Fisher's z method of averaging correlation coefficients.

Judges 1 - 2 = .83
Judges 1 - 3 = .73
Judges 2 - 3 = .72
Day 1 = .73
Day 2 = .81
Day 3 = .78
Comp. = .73

T_1 = Archery
T_2 = Body Mechanics
T_3 = Beginning Swimming
T_4 = Tennis
T_5 = Track and Field

Table 4

Percentage of Agreement Between Judges Rating Frequency of Occurrence of 35 Behavioral Items

	Judges 1-2		Judges 1-3		Judges 2-3		Totals	
	f	%	f	%	f	%	f	%
Clarity and Knowledge of Subject								
1. Verbally presents skills concisely	16	80	14	70	16	80	46	77
2. Verbally presents skills accurately	18	90	15	75	15	75	48	80
3. Verbally presents rules accurately	11	55	15	75	16	80	42	70
4. Verbally interprets rules accurately	16	80	15	75	19	95	50	83
5. Demonstrates skills well	14	70	15	75	17	85	46	77
6. Answers questions about the activity promptly	17	85	8	40	9	45	34	57
7. Answers questions about the activity accurately	19	95	9	45	8	40	36	60
8. Speaks loudly enough for all students to hear	19	95	15	75	14	70	48	80
9. In class, gives directions clearly enough that students follow without question or confusion	14	70	7	35	5	25	26	43
10. Gets students actively involved in learning early in the lesson	17	85	13	65	14	70	44	73
11. Uses a variety of drills or learning experiences	13	65	7	35	12	60	32	53
12. Analyzes individual student errors and tells appropriate corrections	14	70	15	75	16	80	45	75
13. Allows students time to practice	18	90	19	95	17	85	54	90
14. Tells students specific objectives for the class	15	75	18	90	17	85	50	83
Friendliness and Interest in Students								
15. Calls students by name	10	50	15	75	11	55	36	60
16. Maintains good eye contact with students	13	65	15	75	12	60	40	67
17. Yields to class members in a discussion or question-answer period	8	40	9	45	5	25	22	37
18. Listens to students and verbally acknowledges their questions or remarks	17	85	10	50	7	35	34	57
19. Talks with students about things other than class	19	95	16	80	15	75	50	83
20. Praises individual students in front of others	14	70	12	60	14	70	40	67

Table 4 (Continued)

	Judges 1-2		Judges 1-3		Judges 2-3		Totals	
	f	%	f	%	f	%	f	%
21. Moves about among the group	13	65	9	45	14	70	36	60
22. Verbally encourages students individually	15	75	13	65	12	80	40	67
23. Participates with the group	13	65	13	65	16	20	42	70
24. Smiles at students	16	80	5	25	4	20	25	42
25. Talks with students before and after class	12	60	11	55	8	40	31	52
26. Criticizes students excessively	19	95	20	100	19	95	58	97
Fairness								
27. Tells students evaluative techniques in advance	14	70	13	65	17	40	44	73
28. Gives individual attention to students	15	75	15	75	15	75	45	75
29. Spends an excessive amount of time with one or a few students	19	95	20	100	19	95	58	97
Enthusiasm and Sense of Humor								
30. Begins class on time	18	90	18	90	18	90	54	90
31. Verbally praises and encourages the group	15	75	13	65	11	55	39	65
32. Induces students to smile or laugh	15	75	6	30	12	60	33	55
33. Laughs at self when appropriate	14	70	9	45	12	60	35	58
34. Laughs with students	17	85	5	25	4	20	26	43
35. After beginning class, uses entire time allotted	16	80	17	85	19	95	52	87

The data collected were further analyzed to determine if a pattern existed in the type of disagreement between observers. Tallies were made for the number of times Judge 1 marked the frequency of occurrence of an item as frequently or always, while Judge 2 described the item as occurring seldom or never. The same procedure was followed for cases in which Judge 1 said frequently or always when Judge 2 said sometimes, and in which Judge 1 said sometimes when Judge 2 said seldom or never. The reverse was done to study the number of times Judge 2 rated the frequency of occurrence of an item higher than Judge 1. The same dual comparisons were made between Judges 1 and 3, and 2 and 3. The findings revealed that Judge 1 rated behaviors as occurring more frequently than did Judges 2 ($f = 31$) and 3 ($f = 117$), and Judge 2 rated behaviors as occurring more frequently than did Judge 3 ($f = 79$). The reader is referred to Table 5 for a summary of the pattern of disagreement.

Validity

The logical validity of the items on the scale constructed in this paper was studied in a preliminary investigation. The reader is referred to Chapter III for a description of the procedures employed.

The operational validity of the scale was studied in the pilot project and was investigated again in the final study. The reader is referred to Chapter III for a description of the study on operational validity in the pilot project.

Table 5

Pattern of Disagreement Between Observers for 700 Paired Ratings

Observers	Frequently/ Never	Frequently/ Sometimes	Sometimes/ Never	Total	Pattern	Frequency
1 / 2	8[a]	59	24	91	1 > 2	31
2 / 1	12[b]	24	24	60		
1 / 3	15	138	19	172	1 > 3	117
3 / 1	10	12	23	55		
2 / 3	14	121	19	144	2 > 3	79
3 / 2	8	27	30	65		

[a]Indicates that 8 times, Judge 1 marked frequently or always while Judge 2 marked seldom or never.

[b]Indicates that 12 times, Judge 2 marked frequently or always while Judge 1 marked seldom or never.

A conference was held with the observers at the conclusion of the observations in the actual study to discuss the operational use of the final scale. All observers expressed the belief that the items which had been separated during the revision of the trial scale were more easily rated. With respect to the three behaviors which had been added to the scale at the suggestion of the panel of experts, the observers indicated that one, "maintains good eye contact with students," was difficult to judge. They further suggested that the words "in learning" be eliminated from the item "gets students actively involved in learning early in the lesson," since they thought it was easy to judge whether or not students were actively involved in the class, but difficult or impossible to judge whether or not learning was taking place.

The scorers expressed the opinion that rating the frequency of occurrence of the behavioral items on the scale was easier when considerable instruction occurred.

As in the pilot study, the judges believed that concentrating so many observations of different teachers in a brief period of time may have impaired their ability to make clear judgments on the composite ratings.

INTERPRETATION OF FINDINGS

Objectivity

A comparison of the resulting correlation coefficient for objectivity of .76 for this study with results reported in earlier related studies indicated that the degree of agreement between

judges was characteristic for a scale of this nature. Other investigators studying objectivity for various descriptive instruments reported coefficients ranging from .60 to .79, and percentages ranging from 65 to 94 (1, 8, 51, 60, 70).

The slight increase from the trial scale to the revised scale in the size of the correlation coefficient for objectivity indicated that the changes made in the items on the form prior to the final study may have eliminated some ambiguity. It was the opinion of the observers and the investigator that the overall coefficient for objectivity, as well as the correlation coefficient for the composite ratings, might have been higher if the observations for all teachers had not been scheduled within a relatively short period of time.

The wide range in the sizes of correlation coefficients for the various classes (.63 - .84) indicated that the scale may be more appropriate for use in some activity classes than in others. The opinion of the observers throughout the study was that the scale would be more easily utilized in team sports classes than in individual activities. Data from the actual study showed that agreement was highest in swimming and tennis classes, and lowest in archery. The investigator knows of no explanation for this finding. No team sports classes were included in the present investigation.

The fact that overall coefficients for Days 2 and 3 were higher than Day 1 may be indicative of a need for observers to

adjust to various teachers and activity settings prior to achieving objectivity in their ratings.

The fact that eight of the coefficients of correlation for objectivity were .90 or higher, and 23 of the coefficients were .80 or above, in addition to the finding that agreement between two of the observers (Judges 1 and 2) was .83 for 20 independent ratings, indicated that high agreement is possible between judges using the scale developed in the present study. This conclusion is further supported by the finding that all judges agreed on the frequency of occurrence of 31 of the 35 items on the scale at least 50 per cent of the time, and on 16 of the items at least 75 per cent of the time.

The wide range in percentage of agreement between paired observers on the frequency of occurrence of the 35 items indicated that, while some items were very easily rated, others were rated alike as little as one-fifth of the time. It is important to note that, although agreement was low between paired observers on a number of items, at least two observers agreed 70 per cent of the time or more on the frequency of occurrence of all except three of the behaviors: (1) uses a variety of drills or learning experiences, (2) yields to class members in a discussion or question-answer period, and (3) talks with students before and after class. The lack of agreement on the first item may have been a result of its relative nature to the teaching process and to different classes. It is the opinion of the investigator that low agreement on the second item was due to lack of clarity

of the word "yields." Poor agreement on the third item may have been the result of lack of adequate opportunity on the part of the judges to observe teachers before and after class, although they were instructed to do so. It is the opinion of the writer that, for future use or study of the scale, serious consideration should be given to eliminating these three items.

The fact that Judge 1 rated behaviors as occurring more frequently than did Judges 2 and 3, and Judge 2 rated items as occurring more often that Judge 3, indicated that the pattern of disagreement between observers was relatively consistent using the categories frequently or always, sometimes, and seldom or never. This indicated that differences existed between observers in their interpretations of the meanings of the three units of analysis. This finding is not totally unexpected since the investigator made no attempt to specifically define the three categories because of the relative nature of each of the 35 items to the teaching process. For example, on item number 30, if a teacher began class on time, that item was rated as occurring frequently or always. In contrast, each observer rated the frequency of occurrence of item number 34, laughs with students, in terms of her own expectations and interpretation of the units of analysis. It is the opinion of the writer that this is an inborn fault of a scale of this nature which would be difficult or impossible to eliminate. Further study is needed to determine whether or not this problem can be remedied.

Testing the hypothesis. The findings in the present study regarding objectivity resulted in the rejection of the null hypothesis. A table on the values of the correlation coefficient for different levels of significance indicated that for 58 degrees of freedom, coefficients of .25 and .33 or greater were necessary for significance at the .05 and .01 levels respectively (13). Since all of the correlation coefficients computed from the data collected in the present study exceeded these standards, the hypothesis that no significant relationship existed between or among the results obtained by independent scorers using the scale under the same circumstances was rejected.

Validity

The raters indicated at the conclusion of the study that the scale constructed herein was operationally valid. Several observations were made which merit discussion.

Although three observations appear to be sufficient prior to making a reliable composite rating, the objectivity of completing the composite form would likely be enhanced by observing only one teacher during any given period of time.

Class periods 50 minutes in length appear to be sufficiently long for marking the frequency of occurrence of the behavioral items on the scale, and longer periods do not seem to improve objectivity. It should be noted that the latter observation is based on very little data from the actual study.

A need exists to use the scale repeatedly in several different types of activity classes to determine if it is

particularly suitable for use in observing teachers in certain kinds of activities. Regardless of the type of class in which observations are being made, however, it is the opinion of the observers and the investigator that the scale is more functional when instruction is occurring.

Finally, for future use, consideration should be given to modifying one of the items on the scale and eliminating another. The words "in learning" in the item "gets students actively involved in learning early in the lesson," should be eliminated in order to obviate ambiguity. Since the observers indicated that one item, maintains good eye contact with students, was difficult to judge, and since the degree of agreement for that item was not particularly high (67 per cent), consideration should be given to eliminating it from further use of the scale.

SUMMARY

The data collected in the present study regarding objectivity were analyzed by means of the Pearson product-moment method of correlation, Fisher's z method for averaging coefficients of correlation, and the percentage of agreement method. The overall correlation coefficient for agreement among judges rating the frequency of occurrence of 35 behavioral items was .76. The range of agreement on individual items was 37 to 97 per cent, with all except four items showing agreement at least 50 per cent of the time. The correlation coefficient was significant at the .01 level, and the hypothesis that no significant

relationship existed between or among results obtained by independent scorers using the scale under the same circumstances was rejected.

In a final interview with the observers at the completion of the collection of data, the raters confirmed the operational validity of the scale. Further discussion indicated that: 50 minutes was sufficient time for marking the scale during an observation; three observations were sufficient prior to completing a composite rating; and, more research is necessary to determine if the scale is particularly functional when observing teachers in certain types of activities.

CHAPTER V

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

The purpose of the present chapter is three-fold: (1) to summarize the present study, (2) to state the results and conclusions of the study, and (3) to make recommendations for further study.

SUMMARY

The following is a brief resume of the first four chapters of the present study. The chapters are reviewed in the order in which they were presented in the paper.

Introduction

Need for the study. Teacher behavior and effectiveness are critical factors in the learning process at all levels of education, yet few facts have been established regarding the nature of these phenomena. Because information regarding teacher behavior is relevant to such critical issues as teacher selection procedures, teaching quality, and administrative decisions on academic rank, salary, tenure, and merit raises, it is important that efforts be made to obtain information regarding the characteristics and behaviors of teachers in very specific situations. Therefore, the writer became interested in studying specific behavioral correlates of characteristics

which have been identified as showing strong relationships to teaching effectiveness, and in developing a conceptual system for identifying teaching behaviors in college physical education activity classes.

Statement of the problem. The study was designed to develop a valid scale which may be used to observe and describe the frequency of occurrence of selected teacher behaviors in physical education activity classes. A related problem was to study the operational validity of the items on the scale. A second related problem was to determine the objectivity of utilizing the scale constructed.

Purpose of the study. The general purpose of the present study was to develop an instrument which may be used to observe and describe teacher behavior in a physical education activity setting. Specific purposes included the following:

1. To identify characteristics of effective teachers.
2. To identify behaviors which are correlates of the characteristics of effective teachers.
3. To construct a scale which may be used to observe and describe the frequency of occurrence of selected teacher behaviors in college physical education activity classes.
4. To establish the validity of the scale constructed in the study.
5. To study the objectivity of rating the frequency of occurrence of the behaviors selected for use on the scale constructed in this study.

Hypothesis. The hypothesis tested in the present study was as follows: regarding the use of the scale constructed in the present study, there is no significant relationship between or among the results obtained by independent scorers using the scale under the same circumstances.

Review of Selected Research

A review of studies reporting over 30,000 opinions of students and teachers regarding qualities necessary to effective teaching revealed seven concepts which have been identified consistently as being critical to the teaching process. They were: enthusiasm, knowledge of subject, fairness, clarity, friendliness, interest in students, and sense of humor. It appeared from the literature, however, that these concepts lacked operational definitions.

Most of the scales which have been developed for use in observing and describing teacher behavior have been highly complex instruments which have focused on student and teacher verbal behavior and overall classroom climate. The general procedure underlying their development has been three-fold: first, finding a basis for, and developing a system of categorization; second, deciding upon a unit of analysis; and finally, determining overall procedures and means of summarizing the observations.

Procedures

Seven qualities of effective teaching were identified through a study of the literature, and behavioral correlates of these concepts were developed through documentary analysis and introspection. The objectivity and validity of a trial instrument were studied during a pilot project by submitting the trial scale to a jury of nine experts for verification, and by utilizing the scale during 60 independent observations on five different teachers.

Results revealed a correlation coefficient of .73 for objectivity, indicating fairly high agreement among observers, and the scale, with six minor revisions, was confirmed as being valid.

For the actual study, three daily ratings and a composite rating on five different teachers were made independently by three observers. On each rating, 35 behavioral items were marked as occurring frequently or always, sometimes, or seldom or never. The data were analyzed by means of the Pearson product-moment method of correlation, Fisher's z method for averaging correlation coefficients, and the percentage of agreement method. Information regarding the operational validity of the final scale was collected, reported and discussed.

Analysis of Data

The statistical techniques employed in determining the objectivity of the scale constructed in the study were the percentage of agreement method, the Pearson product-moment method

of correlation, and Fisher's z method for averaging correlation coefficients. The objectivity for a total of 60 observations recorded by independent raters was .76. The percentage of agreement for each item was 50 per cent or higher on 31 of the 35 behaviors, and 75 per cent or better on 16 of the 35 items. Overall results were sufficiently high (.01 level of significance) for rejecting the hypothesis that no significant relationship existed between or among results obtained by independent scorers using the scale under the same circumstances.

The observers who had practiced using the scale confirmed the operational validity of the instrument. Related findings indicated that: 50 minutes was a sufficiently long period of time to obtain a reliable rating; three observations were enough prior to completing a composite rating; and, different types of activity settings may have varying effects on the objectivity and operational validity of the scale.

CONCLUSIONS

The following conclusions are based upon the data collected in the present study:

1. The objectivity of judging the frequency of occurrence of the items on the scale was fairly high, as indicated by a correlation coefficient of .76.
2. The objectivity for most of the individual items

was reasonably good, as indicated by the fact that agreement on 31 of the 35 items was 50 per cent or higher, and agreement on 16 of the items, 75 per cent or higher.

3. The scale constructed in this study is logically and operationally valid for use in observing and describing the frequency of occurrence of selected teacher behaviors in college physical education activity classes.

RECOMMENDATIONS

During the conduct of the present investigation, several problems related to the study emerged which merit further research. Appropriate investigation might include:

1. Repeating the study with a larger number of observers.
2. Repeating the study with a larger number of teachers.
3. Repeating the study in different types of activity classes, including team sports, dual and individual sports, and dance.
4. Repeating the study on a larger scale but limiting the raters to observing only one teacher during any given period of time.
5. Studying the appropriateness of the scale for use in observing physical education teachers on the elementary and secondary levels.

6. Studying other units of analysis to use with the scale.
7. Studying the evaluative, rather than descriptive, worth of the scale constructed in this study.

BIBLIOGRAPHY

BIBLIOGRAPHY

A. BOOKS

1. Beecher, Dwight E. The Evaluation of Teaching: Backgrounds and Concepts. New York: Syracuse University Press, 1949.

2. Biddle, Bruce J., and William J. Ellena (eds.). Contemporary Research on Teacher Effectiveness. New York: Holt, Rinehart and Winston, Inc., 1964.

3. Brown, James W., and James W. Thornton, Jr. College Teaching: Perspectives and Guidelines. New York: McGraw-Hill Book Company, Inc., 1963.

4. Cole, Luella. The Background for College Teaching. New York: Farrar and Rinehart, Inc., 1940.

5. Domas, Simeon J. Report on an Exploratory Study of Teacher Competence. Cambridge, Massachusetts: The New England School Development Council, 1950.

6. Flanders, Ned A. Analyzing Teacher Behavior. Reading, Massachusetts: Addison-Wesley Publishing Company, 1970.

7. Guilford, J. P. Fundamental Statistics in Psychology and Education. New York: McGraw-Hill Book Company, 1965.

8. Hyman, Ronald T. (ed.). Teaching: Vantage Points for Study. New York: J. B. Lippincott Company, 1968.

9. Mathews, Donald K. Measurement in Physical Education. Philadelphia: W. B. Saunders Company, 1968.

10. Ryans, David G. Characteristics of Teachers. Washington, D. C.: American Council on Education, 1960.

11. Smith, B. Othanel (ed.). Research in Teacher Education: A Symposium. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1971.

12. Walker, Helen M., and Joseph Lev. Elementary Statistical Methods. New York: Holt, Rinehart and Winston, 1958.

13. Weber, Jerome C., and David R. Lamb. Statistics and Research in Physical Education. St. Louis: The C. V. Mosby Company, 1970.

B. PERIODICALS

14. Astin, Alexander W., and Calvin B. T. Lee. "Current Practices in the Evaluation and Training of College Teachers," Educational Record, XLVII (Summer, 1966), 361-75.

15. Bannister, John, and others. "Evaluating College Teaching," Curriculum Reporter Supplement No. 1, (December, 1961), 1-8.

16. Barr, Arvil S. "The Measurement of Teacher Characteristics and Prediction of Teaching Efficiency," Review of Educational Research, XXII (June, 1952), 169-74.

17. Barr, A. S. "Teacher Effectiveness and Its Correlates," Journal of Experimental Education, XXX (September, 1961), 134-56.

18. Bayley, David H. "Making College Teaching a Profession," Improving College and University Teaching, XV (Spring, 1967), 115-19.

19. Behling, Mary Alice. "The Development of a Screening Program for the Selection and Retention of Women Physical Education Major Students," Dissertation Abstracts, XXX (April, 1970), 4258-59A.

20. Belknap, Robert H., and others. "Guidelines for Promotion," Improving College and University Teaching, XIII (Winter, 1965), 14-15.

21. Bousfield, W. A. "Students' Ratings of Qualities Considered Desirable in College Professors," School and Society, LI (February, 1940), 253-56.

22. Boyer, Marcia. "Teacher Evaluation: Toward Improving Instruction," Junior College Research Review, IV (January, 1970), 1-4.

23. Chiu, Lian-Hwang. "Application of Self-Anchoring Scaling for Study of Teaching Effectiveness," Journal of Educational Research, LXV (March, 1972), 317-20.

24. Combs, Arthur W. "The Personal Approach to Good Teaching," Educational Leadership, XXI (March, 1964), 369-77, 399.

25. deBriun, Hendrik C. "Quality Instruction," Improving College and University Teaching, XV (Autumn, 1967), 214-15.

26. Eckert, Ruth E. "Ways of Evaluating College Teaching," School and Society, LXX (February, 1950), 65-69.

27. Finn, Stanley R. (ed.). "Characteristics of the Effective College Teacher," Higher Education Research Abstracts No. 8, (September, 1969), 1-8.

28. Flanders, Ned A. "Teacher Influence, Pupil Attitudes, and Achievement," Cooperative Research Monograph No. 12. Washington, D. C.: Department of Health Education, and Welfare, Office of Education, 1965.

29. Gage, N. L. "The Appraisal of College Teaching," The Journal of Higher Education, XXXII (January, 1961), 17-22.

30. Galloway, Charles M. "Teacher Nonverbal Communication," Educational Leadership, XXIV (October, 1966), 55-63.

31. Gibb, Cecil A. "Classroom Behavior of the College Teacher," Educational and Psychological Measurement, XV (Autumn, 1955), 254-63.

32. Hoffmann, Randall W. "Students Portray the Excellent Teacher," Improving College and University Teaching, XI (Winter, 1963), 21-24.

33. Isaacson, Robert L., Wilbert J. McKeachie, and John E. Milholland. "Correlation of Teacher Personality Variables and Student Ratings, Journal of Educational Psychology, LIV (April, 1963), 110-17.

34. LaGrand, Louis E. "Better Teaching Through Student Evaluation," The Physical Educator, XXVIII (December, 1971), 201-202.

35. Langen, Thomas D. F. "Student Assessment of Teaching Effectiveness," Improving College and University Teaching, XIV (Winter, 1966), 22-25.

36. Lewis, W. W., J. M. Newell, and J. Withall. "An Analysis of Classroom Patterns of Communication," Psychological Report, IX (July-December, 1961), 211-19.

37. MacDonald, Marion E. "Students' Opinion as Regards Desirable and Undesirable Qualifications and Practices of Their Teachers in Teacher-Training Institutions," Educational Administration and Supervision, XVII (February, 1931), 139-46.

38. Mackey, Helen T. "Job Analysis of Women Supervisors of Physical Education in United States Public Schools," Research Quarterly, XXVII (March, 1956), 32-40.

39. Maslow, A. H., and W. Zimmerman. "College Teaching Ability, Scholarly Activity and Personality," The Journal of Educational Psychology, XLVII (March, 1956), 185-89

40. McComas, J. D. "Profiles of Teachers," Improving College and University Teaching, XIII (Summer, 1965), 135-36.

41. McGrath, Earl J. "Superior Teaching in the Colleges of the Sixties," Improving College and University Teaching, XIII (Summer, 1960), 76-83.

42. Means, Richard K. "A Teacher Appraisal Scale," Journal of Health, Physical Education and Recreation, XXI (May-June, 1960), 36-37.

43. Medley, Donald M., and Harold E. Mitzel. "Some Behavioral Correlates of Teacher Effectiveness," Journal of Educational Psychology, L (December, 1959), 239-46.

44. Morton, Richard K. "Students' Views of Teaching," Improving College and University Teaching, XIII (Summer, 1965), 140-42.

45. Musella, Donald, and Reuben Rusch. "Student Opinion on College Teaching," Improving College and University Teaching, XVI (Spring, 1968), 137-40.

46. Phillips, Beeman N. "The 'Individual' and the 'Classroom Group' as Frames of Reference in Determining Teacher Effectiveness," Journal of Educational Research, LXIII (November, 1964), 128-31.

47. Quick, Alan F., and Arnold D. Wolfe. "The Ideal Professor," Improving College and University Teaching, XIII (Summer, 1965), 133-34.

48. Renner, Richard R. "A Successful Rating Scale," Improving College and University Teaching, XV (Winter, 1967), 12-14.

49. Rosenshine, Barak. "Enthusiastic Teaching: A Research Review," School Review, LXXVIII (August, 1970), 499-514.

50. Rosenshine, Barak. "Evaluation of Classroom Interaction," Review of Educational Research, XL (April, 1970), 279-300.

51. Ryans, David G. "Some Relationships Between Pupil Behavior and Certain Teacher Characteristics," Journal of Educational Psychology, LCC (April, 1961), 82-90.

52. Ryans, David G. "Teacher Behavior Theory and Research: Implications for Teacher Education," Journal of Teacher Education, XIV (September, 1963), 274-93.

53. Schaffle, A. E. F. "The Pupil Checks the Teacher," School Executive, LI (December, 1931), 151-53.

54. Smith, Othanel. "A Conceptual Analysis of Instructional Behavior," Journal of Teacher Education, XIV (September, 1963), 294-98.

55. Solomon, Daniel, Larry Rosenberg, and William E. Bedzek. "Teacher Behavior and Student Learning," Journal of Educational Psychology, LV (February, 1964), 23-30.

56. Solomon, Daniel. "Teacher Behavior Dimensions, Course Characteristics, and Student Evaluations of Teachers," American Educational Research Journal, III (January, 1966), 35-47.

57. Solomon, Daniel, William E. Bedzek, and Larry Rosenberg. "Dimensions of Teacher Behavior," Journal of Experimental Education, XXXIII (Fall, 1964), 23-40.

58. Weaver, Carl H. "Instructor Rating by College Students," Journal of Educational Psychology, LI (February, 1960), 21-25.

59. Williams, Howard Yolen, Jr. "College Students' Perceptions of the Personal Traits and Instructional Procedures of Good and Poor Teachers," Dissertation Abstracts, XXVII (May, 1967), 3644A.

60. Withall, John. "The Development of a Technique for the Measurement of Social-Emotional Climate in Classrooms, Journal of Experimental Education, XVII (March, 1949), 347-61.

61. Yourglich, Anita. "Study on Correlations Between College Teachers, and Students' Concepts of 'Ideal-Student' and 'Ideal-Teacher'," Journal of Educational Research, XLIX (September, 1955), 59-64.

C. UNPUBLISHED MATERIALS

62. Cohen, Arthur M., and Florence B. Brawer. "Measuring Faculty Performance." Washington, D. C.: American Association of Junior Colleges, 1969.

63. Harristhal, Joann Winifred. "A Student Reaction Inventory for Rating Teachers in the College Women's Physical Education Service Program." Unpublished Doctor's dissertation, University of Oregon, 1962.

64. Harsh, J. Richard. "A Look at Teacher Evaluation." Paper read at the Annual Conference of the American Association of School Personnel Administrators, October 22, 1970, Las Vegas, Nevada.

65. Hildebrand, Milton, and Robert C. Wilson. "Effective University Teaching and Its Evaluation." Berkeley: Center for Research and Development in Higher Education, University of California, 1970.

66. Hoyt, Donald P. "Identifying Effective Teaching Behaviors." Manhattan, Kansas: Kansas State University, 1969.

67. Krupka, John G. "Report on Faculty and Student Evaluation of Instructor Rating Questionnaire." Northampton County Area Community College, March, 1970.

68. Morsh, Joseph E., and Eleanor W. Wilder. "Identifying the Effective Instructor: A Review of the Quantitative Studies, 1900-1952." Illinois: Air Force Personnel and Training Research Center, Chanute AFB, 1954.

69. Overturf, C. L., and Edwin C. Price. "Student Rating of Faculty at St. Johns River Junior College." Palatka, Florida: A Study Proposed under Kellogg Grant to Junior College Division of Universities of Florida, St. Johns River Junior College, 1966.

70. Smith, B. Othanel, and Milton Meux. "A Study of the Logic of Teaching." Urbana: University of Illinois, 1962.

APPENDICES

APPENDIX A

Instructions to, and Data from Jury, of Experts

JURY OF EXPERTS

Institutions Represented:

Kent State University	Kent, Ohio
Northwestern State University	Natchitoches, Louisiana
Oklahoma State University	Stillwater, Oklahoma
Southern Illinois University	Carbondale, Illinois
University of North Carolina	Greensboro, North Carolina

Jury Members	Institutions	Area of Expertise
Dr. Betty Abercrombie	O.S.U.	Teacher Supervisor
Dr. Robert Alost	N.S.U.	Administration
Dr. Kate Barrett	U.N.C.	Teacher Behavior
Dr. John Bayless	O.S.U.	Teacher Supervisor
Dr. Rosemary McGee	U.N.C.	Scale Construction
Dr. Matthew Resick	K.S.U.	Teacher Behavior
Dr. Beverly Seidel	K.S.U.	Administration
Dr. Celeste Ulrich	U.N.C.	Behavioral Bases of Physical Education
Dr. Charlotte West	S.I.U.	Scale Construction

INSTRUCTIONS TO JURY EXPERTS

The purpose of the present investigation is to develop a conceptual system for identifying teaching behaviors in physical education activity classes on the college level.

After a survey of the related literature, seven concepts which have been identified consistently as characterizing effective teachers were chosen for the purpose of this study. They are: clarity, enthusiasm, fairness, friendliness, interest in students, knowledge of subject, and sense of humor. Introspection and further analysis of the research yielded behavioral correlates of each of these concepts.

At present, a pilot study is being conducted to determine the objectivity of a trial instrument on which the behaviors have been classified by the investigator. For each of five different instructors teaching various activities, several raters are observing three consecutive class periods and marking independently the frequency with which each behavior occurs. Observers are rating frequency according to three categories: (1) frequently or always, (2) sometimes, and (3) seldom or never. The objectivity of the scale will be studied for Day 1, Day 2, Day 3, and for a composite rating completed at the end of Day 3. The composite rating will be filled out independent of the other ratings. It should be emphasized that the observers are concerned only with describing how often behaviors occur, not with valuing whether the frequency or infrequency is good or bad.

You are one of a jury of ten experts in the areas of teacher behavior, scale construction, the behavioral bases of physical education, administration of physical education, and supervision of student teachers in physical education who have been selected to aid in validation of the proposed scale.

On the enclosed form, the behavioral items chosen for the present study have been randomized in order to obviate possible patterning in responses. Please use an "X" to indicate the one concept to which you think each behavioral correlate belongs. If you think an item is not behaviorally stated, or is not reflective of any of the concepts, please check the "NO" column. In addition, please check whether you think each item is or is not observable in a physical education activity class.

After the comments and suggestions of all jury members are received, the scale will be subjected to further analysis and possible revision. Finally, the objectivity of utilizing the refined instrument will be studied.

For your convenience, a self-addressed, stamped envelope is enclosed in which you may return the information. Your assistance with this project is deeply appreciated.

Judy Showers

COPY OF SCALE USED BY JURY MEMBERS

Jury Member _____

Date _____

	Clarity	Enthu-siasm	Fair-ness	Friend-liness	Interest in Students	Knowledge of Subject	Sense of Humor	NO	Is the Item Observable?	
									Yes	No
1. Verbally presents skills concisely and accurately										
2. Laughs with students										
3. Allows students to tell jokes or play practical jokes										
4. Uses a variety of drills or learning experiences										
5. Listens to students and verbally acknowledges their questions or remarks										
6. In class, gives directions clearly enough that students follow without question or confusion										
7. Allows students time to practice										
8. Talks with students before and after class										
9. Tells students evaluative techniques in advance										
10. Changes voice inflection										
11. Induces students to smile or laugh										

	Clarity	Enthu-siasm	Fair-ness	Friend-liness	Interest in Students	Knowledge of Subject	Sense of Humor	NO	Is the Item Observable? Yes	No
12. Verbally praises and encourages the group										
13. Demonstrates skills well										
14. Answers questions about the activity promptly and accurately										
15. Yields to class members in a discussion or question-answer period										
16. Dresses appropriately for the activity										
17. Praises individual students in front of others										
18. Participates with the group										
19. Avoids excessive criticism of students										
20. Analyzes individual student errors and tells appropriate corrections										
21. Begins class on time and uses entire time allotted										
22. Gives individual attention to all students and avoids spending an excessive amount of time with one or a few										
23. Speaks loudly enough for all students to hear										

	Clarity	Enthu-siasm	Fair-ness	Friend-liness	Interest in Students	Knowledge of Subject	Sense of Humor	NO	Is the Item Observable?	
									Yes	No
24. Uses gestures										
25. Smiles at students										
26. Verbally encourages each student individually										
27. Moves about among the group										
28. Tells students specific objectives for the class										
29. Verbally presents and interprets rules accurately										
30. Talks with students about things other than class										
31. Calls students by name										

SUGGESTIONS FOR ADDITIONAL BEHAVIOR

COMMENTS OR SUGGESTIONS:

SUMMARY OF CLASSIFICATION OF BEHAVIORAL ITEMS
BY JURY OF EXPERTS

Jury Member _____

Date _____

	Clarity	Enthu-siasm	Fair-ness	Friend-liness	Interest in Students	Knowledge of Subject	Sense of Humor	NO
1. Verbally presents skills concisely and accurately	5			1		2		
2. Laughs with students				2			6	
3. Allows students to tell jokes or play practical jokes					1		4	4
4. Uses a variety of drills or learning experiences	1					6		1
5. Listens to students and verbally acknowledges their questions or remarks			1		8			
6. In class, gives directions clearly enough that students follow without question or confusion	9							
7. Allows students time to practice			2		2	3		2
8. Talks with students before and after class				4	5			
9. Tells students evaluative techniques in advance			7					
10. Changes voice inflection	2	3						4
11. Induces students to smile or laugh		1		1	2		3	1

	Clarity	Enthu-siasm	Fair-ness	Friend-liness	Interest in Students	Knowledge of Subject	Sense of Humor	NO
12. Verbally praises and encourages the group		4			4	1		
13. Demonstrates skills well	2					6		1
14. Answers questions about the activity promptly and accurately			1		1	7		
15. Yields to class members in a discussion or question-answer period			2		6			1
16. Dresses appropriately for the activity		4				1		4
17. Praises individual students in front of others		1			8			
18. Participates with the group		2		4				1
19. Avoids excessive criticism of students			3		6			
20. Analyzes individual student errors and tells appropriate corrections	1					6		1
21. Begins class on time and uses entire time allotted	1	4	4					
22. Gives individual attention to all students and avoids spending an excessive amount of time with one or a few			7		2			

	Clarity	Enthu-siasm	Fair-ness	Friend-liness	Interest in Students	Knowledge of Subject	Sense of Humor	NO
23. Speaks loudly enough for all students to hear	8							1
24. Uses gestures		4						5
25. Smiles at students				7	1			
26. Verbally encourages each student individually		1			8			
27. Moves about among the group					5	1		2
28. Tells students specific objectives for the class	3		2		1	2		
29. Verbally presents and interprets rules accurately	1		1			7		
30. Talks with students about things other than class				2	6			1
31. Calls students by name				3	6			

APPENDIX B

Instructions to Observers and Teachers

MEMORANDUM TO OBSERVERS: PILOT STUDY

Please read the items carefully before beginning the observations. It is important that you be present at the activity area at least five minutes before class is scheduled to begin, and that you remain until after the students have been dismissed.

You will observe each of five different teachers during three consecutive instructional periods. For each day, you are to complete a separate form. The first three are not accumulative; you are to mark the frequency of occurrence of each behavior only for the class period being observed. The final composite rating should be completed at the end of the third day and is an overall rating of the frequency of occurrence of each item observed over the three days. It should be completed without reference to the forms from the first three days.

The purpose of the scale is to describe teacher behavior, not to evaluate or judge it. Please check each item only with respect to the frequency of its occurrence by placing an "X" in one of the three categories.

The following is a schedule of the classes to be observed:

1:30	MWF	Modern Dance	(Dance Studio)
2:30	MWF	Beginning Swimming	(Pool)
8:30	WFM	Adv. Swim. & Lifesav.	(Pool)
12:30	WFM	Body Mechanics	(Body Mechanics Room)
1:30	MWF	Gymnastics	(Wrestling Room)

You should begin observing the first two classes on Monday, January 22. The second two classes you may begin observing on Wednesday, January 24, and the observations of the gymnastics class will begin on Monday, January 29.

Please turn in each rating sheet as soon after you complete it as possible. Your assistance in this project is deeply appreciated.

Judy Showers

MEMORANDUM TO TEACHERS: PILOT STUDY

This is to remind you that three observers will be attending your classes for a week in conjunction with a pilot study I am conducting. They will need to position themselves so that they can both see and hear what is occurring in the class, but will try to be as inconspicuous as possible.

The purpose of the observations is to determine if independent observers agree on the frequency of occurrence of certain teacher behaviors in various activity classes.

Observers will begin on the following dates and will attend each class during three consecutive class sessions. Observations for Modern Dance and Beginning Swimming will begin on Monday, January 22. For Body Mechanics and Advanced Swimming and Lifesaving, they will begin on Wednesday, January 24, and for Gymnastics, on Monday, January 29. Again, it is important that these be instructional sessions. If you decide to give an hour examination, show a long film, or become ill and need a substitute teacher, please notify me as soon as possible. Your cooperation in this project is deeply appreciated.

Judy Showers

MEMORANDUM TO OBSERVERS: ACTUAL STUDY

Please read the items carefully before beginning the first observation. It is important that you be present at the activity area at least five minutes before class is scheduled to begin, and that you remain until after the students have been dismissed.

You will observe each of five different instructors during three consecutive instructional periods. For each day, you are to complete a separate form. The first three are not accumulative; you are to mark the frequency of occurrence of each behavior only for the class period being observed. The final composite rating should be completed at the end of the third day and is an overall rating of the frequency of occurrence of each item observed over the three days. It should be completed without reference to the forms from the first three days.

The following is a schedule of the classes to be observed:

Hour	Days	Class	Begin
2:30	MWM	Track and Field (Track)	M 2/26
1:30	MWF	Body Mechanics (B.M. Rm.)	M 3/5
8:30	TThF	Tennis (G-4)	Tu 3/6
2:30	WFM	Beg. Swimming (Pool)	W 3/7
12:30	MWF	Archery (Archery Range)	M 3/12

Monday, February 26 - 2:30
Wednesday, February 28 - 2:30
Monday, March 5 - 1:30, 2:30
Tuesday, March 6 - 8:30
Wednesday, March 7 - 1:30, 2:30
Thursday, March 8 - 8:30
Friday, March 9 - 8:30, 1:30, 2:30
Monday, March 12 - 12:30, 2:30
Wednesday, March 14 - 12:30
Friday, March 16 - 12:30

Judy Showers

MEMORANDUM TO TEACHERS: ACTUAL STUDY

This is to remind you that three observers will be attending your classes for a week in conjunction with a study I am conducting for my dissertation. They will try to be as inconspicuous as possible, but will need to position themselves so that they can see and hear what is occurring in the class.

Again, the purpose of the study is to determine if independent observers see certain behaviors occurring with the same degree of frequency. The forms they are using are not evaluative, and the information will be kept confidential.

Since it is important that the observations occur during three consecutive instructional class sessions, please let me know if you decide to show a film, give an hour exam, or if you become ill and need a substitute teacher.

Judy Showers

APPENDIX C

Scales Constructed and Used in the Present Study

TRIAL SCALE

Class _____ Teacher _____

Observer # _____ Day # _____ Date _____

	Frequently or Always	Sometimes	Seldom or Never
CLARITY			
1. In class, gives directions clearly enough that students follow without question or confusion			
2. Speaks loudly enough for all students to hear			
3. Tells students specific objectives for the class			
ENTHUSIASM			
4. Changes voice inflection			
5. Verbally praises and encourages the group			
6. Dresses appropriately for the activity			
7. Participates with the group			
8. Begins class on time and uses entire time allotted			
9. Uses gestures			
FAIRNESS			
10. Listens to students and verbally acknowledges their questions or remarks			
11. Allows students time to practice			
12. Tells students evaluative techniques in advance			
13. Avoids excessive criticism of students			
14. Gives individual attention to all students and avoids spending an excessive amount of time with one or a few			

(continued)

	Frequently or Always	Some-times	Seldom or Never
FRIENDLINESS			
15. Laughs with students			
16. Smiles at students			
17. Talks with students about things other than class			
INTEREST IN STUDENTS			
18. Talks with students before and after class			
19. Yields to class members in a discussion or question-answer period			
20. Praises individual students in front of others			
21. Verbally encourages each student individually			
22. Moves about among the group			
23. Calls students by name			
KNOWLEDGE OF SUBJECT			
24. Verbally presents skills concisely and accurately			
25. Uses a variety of drills or learning experiences			
26. Demonstrates skills well			
27. Answers questions about the activity promptly and accurately			
28. Analyzes individual student errors and tells appropriate corrections			
29. Verbally presents and interprets rules accurately			
SENSE OF HUMOR			
30. Allows students to tell jokes or play practical jokes			
31. Induces students to smile or laugh			

FINAL SCALE

Class _____ Teacher _____

Observer # _____ Day # _____ Date _____

	Frequently or Always	Some-times	Seldom or Never
CLARITY AND KNOWLEDGE OF SUBJECT			
1. Verbally presents skills concisely			
2. Verbally presents skills accurately			
3. Verbally presents rules accurately			
4. Verbally interprets rules accurately			
5. Demonstrates skills well			
6. Answers questions about the activity promptly			
7. Answers questions about the activity accurately			
8. Speaks loudly enough for all students to hear			
9. In class, gives directions clearly enough that students follow without question or confusion			
10. Gets students actively involved in learning early in the lesson			
11. Uses a variety of drills or learning experiences			
12. Analyzes individual student errors and tells appropriate corrections			
13. Allows students time to practice			
14. Tells students specific objectives for the class			
FRIENDLINESS AND INTEREST IN STUDENTS			
15. Calls students by name			
16. Maintains good eye contact with students			
17. Yields to class members in a discussion or question-answer period			

(continued)

	Frequently or Always	Some-times	Seldom or Never
18. Listens to students and verbally acknowledges their questions or remarks			
19. Talks with students about things other than class			
20. Praises individual students in front of others			
21. Moves about among the group			
22. Verbally encourages students individually			
23. Participates with the group			
24. Smiles at students			
25. Talks with students before and after class			
26. Criticizes students excessively			
FAIRNESS			
27. Tells students evaluative techniques in advance			
28. Gives individual attention to students			
29. Spends an excessive amount of time with one or a few students			
ENTHUSIASM AND SENSE OF HUMOR			
30. Begins class on time			
31. Verbally praises and encourages the group			
32. Induces students to smile or laugh			
33. Laughs at self when appropriate			
34. Laughs with students			
35. After beginning class, uses entire time allotted			