DE GENARO, Arthur Paul, 1933-
EXPERIMENTAL USE OF THE VIDEO TAPE
RECORDER AS AN EVALUATIVE INSTRUMENT
AND OBSERVATIONAL TOOL IN THE
SUPERVISION OF STUDENT TEACHERS OF
PHYSICAL EDUCATION.

The Ohio State University, Ph.D., 1969
Education, teacher training
University Microfilms, Inc., Ann Arbor, Michigan
EXPERIMENTAL USE OF THE VIDEO TAPE RECORDER AS AN 
EVALUATIVE INSTRUMENT AND OBSERVATIONAL TOOL IN 
THE SUPERVISION OF STUDENT TEACHERS OF 
PHYSICAL EDUCATION 

DISSERTATION 
Presented in Partial Fulfillment of the Requirements for 
the Degree Doctor of Philosophy in the Graduate 
School of The Ohio State University 

By 
Arthur Paul De Genaro, B.S., M.Ed. 

* * * * * * 

The Ohio State University 
1969 

Approved by 
Lewis A. Hess 
Advisor
ACKNOWLEDGMENTS

This writer wishes to thank Dr. Lewis A. Hess for his guidance and help in the preparation and development of this paper. Additional thanks and appreciation are extended to Dr. Chalmer G. Hixson and Dr. Charles L. Mand for their support throughout the experiment. Sincere thanks are extended to Dr. Edward Coates and Dr. John Hendrix who served as raters during the course of this experiment.

Thanks must be given to the six student teachers who permitted themselves to be taped as subjects in the study. Without them the this study could not have been possible.

A special note of appreciation is extended to Dr. Allen E. Koenig, Editor of the Educational Broadcasting Review. His valuable professional advice and direction in securing sources for this study was extremely helpful.
VITA

February 11, 1933  Born - Youngstown, Ohio
1955  . . . . . .  B.S., Kent State University, Kent, Ohio
1955-1957  . . .  Officer U.S. Marine Corps
1957-1959  . . .  Teacher Coach, Ravenna City High School, Ravenna, Ohio
1959-1962  . . .  Loan Officer, Portage County National Bank, Kent, Ohio
1963  . . . . . .  M.Ed., Kent State University, Kent, Ohio
1963-1965  . . .  Instructor, Health and Physical Education, Department, Kent State University, Ashtabula Academic Center, Ashtabula, Ohio
1964-1965  . . .  Assistant Director, Kent State University Ashtabula Academic Center, Ashtabula, Ohio. Appointee, Director Geauga Center, Kent State Branch Campus, Chesterland, Ohio
1966-1967  . . .  Research Assistant, Higher Education, The Ohio State University, Columbus, Ohio
1967-1968  . . .  Teaching Associate, Division of Men's Physical Education, The Ohio State University, Columbus, Ohio
1968-1969  . . .  Instructor, Division of Men's Physical Education, The Ohio State University, Columbus, Ohio
VITA (Contd.)

FIELDS OF STUDY

Major Field: Physical Education

Professors Lewis A. Hess, Seymour Kleinman, Margaret A. Mordy, Virginia Craft, Chalmer G. Hixson

Minor Field: Higher Education

Professors Richard J. Frankie, Robert B. Sutton, Everett J. Kircher
TABLE OF CONTENTS

ACKNOWLEDGMENTS .................................................. ii
VITA ................................................................. iii
LIST OF TABLES ..................................................... vii
LIST OF FIGURES .................................................. viii

Chapter

I. INTRODUCTION ................................................... 1
   Statement of the Problem
   Need for the Study
   Definition of Terms
   Delimitations
   Limitations
   Null Hypotheses

II. REVIEW OF LITERATURE ......................................... 11
   Relative Effectiveness Studies
   Non-significant Differences
   Observational Studies of Demonstration Teaching
   Summary

III. METHODS AND PROCEDURES ................................... 33
   Phase One
   Phase Two
   Treatment of the Data

IV. ANALYSIS OF THE DATA ........................................ 41

V. SUMMARY, CONCLUSIONS AND RECOMMENDATIONS ............. 51
TABLE OF CONTENTS (Contd.)

APPENDIX

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A.</td>
<td>57</td>
</tr>
<tr>
<td>B.</td>
<td>60</td>
</tr>
<tr>
<td>C.</td>
<td>62</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>64</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table                             Page
1. Summary of Hotelling $T^2$ Scores among Non-medium and Medium Raters with the Difference Equal to Zero  43
2. Summary of Paired $t$ Difference between Non-medium and Medium Rating Groups  49
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Observation Time of Experimental and Control Groups</td>
<td>19</td>
</tr>
<tr>
<td>2.</td>
<td>Hotelling $T^2$ Acceptance and Rejection Areas</td>
<td>44</td>
</tr>
<tr>
<td>3.</td>
<td>Parid Acceptance and Rejection Areas</td>
<td>46</td>
</tr>
</tbody>
</table>
CHAPTER I

INTRODUCTION

One of the concerns that is paramount today in the area of higher education is that of cost. Within any university the problem of faculty and staff salaries is commensurate with this concern. Corson comments:

Most American colleges and universities exist in relative poverty. Their earned income meets only a fraction of their annual operating cost. Not more than 10 per cent of all private and public colleges and universities have endowment income sufficient to provide as much as one-fifth of their annual budget.¹

The relevancy of the above statement is reflected by the President's report of The Ohio State University 1968, where from total expenditures of $194,696,115.00, some $66,151,993.00 were channeled into Departmental Instruction (Faculty and Staff Salaries).² This represents the largest expenditure of all the activities supported by the University, or some 34 per cent of the total.

In the face of ever increasing expenditures for faculty and staff salaries, institutions with teacher education programs are concerned with the problem of maintaining


²Columbus Dispatch Supplement, September 29, 1968, p. 15.
important phases in the professional preparation of teachers. The field experience program is one area in the preparation of teachers that is generally accepted as being a worthwhile experience.

Educational critics, such as Conant, agree that the field experience program is an important and meaningful facet in the preparation of teachers. Institutions preparing teachers are responsible for the organization and administration of these programs. The supervision and observation of student teachers in laboratory field experience programs, creates a problem of providing sufficient supervisory staff for such a program. Normally faculty members are not able to devote full time to such responsibilities. Along with being assigned a certain number of student teachers, many faculty involved with the supervisory aspect of teacher education have a heavy load of other responsibilities. These responsibilities range from teaching both undergraduate and graduate courses, counseling and advising graduate students, to heading committees, sections and individual research projects; along with participation in organizations within the profession. In addition to full teaching schedules and varied staff responsibilities the supervisor has been limited by the geographic location of the school to be visited. These factors have been a hinderance in performing the necessary observations. In addition to observation, the university supervisor's duties and responsibilities extend
to other supervisory functions. One is counseling the student teacher as he progresses in his field experience. This often is accomplished by means of conferences and seminars. Another is to meet with the cooperating teacher for a mutual exchange of information about their student teacher. The university supervisor is ultimately responsible for the final evaluation of the student teacher whether this evaluation takes the form of a letter grade, a written paragraph, or perhaps a satisfactory pass or non-pass marking. As a result of these many responsibilities, coupled with the previously mentioned factors of time and distance, the supervisor's visits are usually infrequent and un-systematic. This problem of frequency was referred to by Wilson when he stated that, "The number of times such observations were made in semester schools ranged from one to nine with the average being two and one-half visits."¹ It is felt, therefore, that the rationale of this study would be to improve the effectiveness of evaluation and supervision of physical education student teachers by university staff personnel involved in teacher education.

Statement of the Problem

The problem was to determine the capability of the Video Tape Recorder as an evaluative instrument and

observational tool in the supervision of physical education student teachers.

**Need for the Study**

The evolution for this study came from experience while supervising student teachers of physical education at The Ohio State University during a number of academic quarters. It was observed that an excessive amount of time was involved in the actual physical visitations. Additional impetus for the study was given when the teacher education section suggested that video tapes be taken of all student teachers. It was suggested that each student teacher be taped at least once during his internship. The rationale for this was in the expected benefits derived from the replay and reviewing quality, conferences and counseling sessions that could be conducted with the student teacher as a result of the previously taken tapes. The tapes were limited in their use to the teaching learning process and the improvement of instruction. They were not used for specific evaluations or as a substitute for physical visits.

The need for greater staff utilization, economy of time and reduction of physical visitations by those involved in supervision opened a new area of research resulting in this study. The question was one of using the Video Tape Recorder as a tool or instrument for evaluation rather than only as a means of instruction.
According to Inlow, the major problems in evaluating the student teacher experiences are:

1. Who should evaluate?
2. Who should be evaluated?
3. How can evaluation be accomplished most effectively?1

It was to this third area of evaluating the student teaching experience that this study addressed itself. The underlying administrative concept was one of increasing the frequency of observations while decreasing the physical visits of the university supervisor in securing such observation. This concept of increased observations and multi-appraisals is referred to by Fishbach when he comments that:

If numerous appraisals are made, final evaluative judgments of a student teacher's success are arrived at much more easily and honestly. The assumption that final evaluation of a student teacher's success can be one of a post-mortem nature based on memory or sketchy notes of progress is not sound.2

If Meier is correct when he suggests that:

A year's internship, liberally salted with micro-training experiences would be superior to present student teaching practices, then the need is for a more effective, less time consuming technique of student teaching evaluation.3

---


A study utilizing Video Tape Recordings would investigate such a longitudinal concept of evaluating student teaching experiences.

The innovation that could develop for evaluating student teachers might take the form of having a technician record by means of video tape, the student teachers on a formalized, systematic basis. These tapes would be readily available for the supervisor and could be viewed at his convenience. Without leaving his office the supervisor could view the tapes as often as necessary while continuing to meet his teaching responsibilities and other staff functions. As the frequency in taping student teachers increases, the concept of evaluation as a continuous process would be enhanced. The student teacher could now be observed more often and under a systematic, controlled and objective process. This does not imply that direct observation is not objective, but that the typical reaction of what student teachers expect has been a rather traditional one; that the supervisor's role was primarily one of instruction. Delaney and Moore suggest that: "The experience of an experimental or didactic nature should be developed so as to give the student teacher a better understanding of how he is evaluated."¹ It is within this area of improving the

¹Daniel J. Delaney and James C. Moore, "Student Expectations of the Role of Practicum Supervisor," Counselor Education and Supervision, VI (Fall, 1966), 11-17.
technique of evaluation that this study seeks to utilize
the Video Tape Recorder as an effective instrument in such
evaluations.

**Definition of Terms**

<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>VTR</td>
<td>Video Tape Recorder—Vehicle that produces the video tapes.</td>
</tr>
<tr>
<td>Medium</td>
<td>The use of Video Tapes for observation.</td>
</tr>
<tr>
<td>Non-Medium</td>
<td>The use of physical or direct observation.</td>
</tr>
<tr>
<td>Subjects</td>
<td>Observed student teachers in teaching sessions.</td>
</tr>
<tr>
<td>Cells</td>
<td>Areas and locations within the design.</td>
</tr>
<tr>
<td>Check List</td>
<td>Constants to be used by both non-medium and medium raters.</td>
</tr>
<tr>
<td>Evaluation</td>
<td>The process of evaluating. In this study evaluation was accomplished by raters viewing subjects either physically (live) or by means of video tape.</td>
</tr>
<tr>
<td>Supervision</td>
<td>The function of evaluation (tools). The functional method in this study was by a series of systematic observations.</td>
</tr>
<tr>
<td>Formal Systematic Observation</td>
<td>This type of observation uses a predetermined units and time samples. When used in a systematic and objective manner these techniques increase reliability and validity.</td>
</tr>
<tr>
<td>Hotelling $T^2$</td>
<td>Statistical treatment for differences among non-medium and medium raters. Also tends to seek out internal agreement among raters.</td>
</tr>
<tr>
<td>Parid $t$</td>
<td>Uses composite scores of Hotelling $T^2$ and seeks symmetry between scores of medium and non-medium rating groups.</td>
</tr>
<tr>
<td>CCTV</td>
<td>Closed-circuit television is a system to transmit signals or images and/or sound to predetermined receiver or monitor locations. The pivotal feature of closed-circuit television is the element of control exercised over its reception and hence, over its</td>
</tr>
</tbody>
</table>
utilization. The rational for CCTV is in the advantages earned by limiting and controlling the signal.

**Delimitations**

The subjects consisted of male student teachers assigned through the teacher education office of the Division of Men's Physical Education, The Ohio State University, Columbus, Ohio. The subjects were observed during the Autumn Quarter of 1968 on the high school and junior high school levels only. The subjects were in a similar stage of professional preparation, having completed the required physical education courses, general and professional education courses, and pre-requisites for student teaching. The subjects had been accepted and were presently engaged in their student teaching assignments. No attempt was made to relate this study to any phase of the teaching-learning process.

**Limitations**

The scope of this experimental research study was extremely limited. Therefore, not all factors present in actual live situations (observations in-person) were able to be presented by means of the Video Tape Recordings. Such factors as the socio-economic conditions of the school and community were not portrayed by the VTR.

The attitude and over-all tenor of the school was not within the capability of the video tape recordings or design
of this study. The study excluded the areas of feed back and attitude between the teachers and pupils. Also the interpersonal relationships and rapport between the cooperating teachers and the student teachers was not of consideration in this study. Additional areas such as physical properties of the school and gymnasium environment were considered as definite limitations to the study. While the visual images and audio sounds were able to be taped, the olfactory and tactile senses were impossible to capture and communicate by video tape.

Whenever possible these items to be rated were identical to both live and taped situations. Inasmuch as the logistical considerations of observation research have been established as one of the most expensive and time consuming types of research, no sub-problems were attempted in this study.1 All raters were members of the Men's Physical Education Staff at the Ohio State University.

**Null Hypotheses**

1. There will be no agreement in the evaluations of the subjects by the three raters observing by direct observation (non-medium ratings).

---

2. There will be no agreement in the evaluations of the subjects by three different raters observing by means of video tapes (medium ratings).

3. There will be no agreement between composite evaluations of subjects observed and rated by direct physical observation (non-medium ratings), and the composite evaluations of those same subjects when observed and rated by means of video tapes (medium ratings).
There appears to be an absence of research in the use of television tapes as a means of supervising and evaluating student teachers. The great majority of studies using this medium were related to specific subject matter areas. These studies were conducted in the setting of a normal or traditional classroom atmosphere. By far the largest category of research was what Greenhill called 'Relative Effectiveness' studies.¹ These were studies in which the performance of students instructed via television had been compared with the performance of others instructed directly, or face-to-face, by a teacher in the usual way. The perponderance of these studies centered around television as a means of instruction. This use of television appeared to be the most important trend in Educational Television Research in the last decade. Chu and Schramm revealed that about half the Educational

Television Research literature involved such studies.¹

A non-significant difference syndrome had been revealed from the bulk of these relative effectiveness studies. Schramm feels that these studies have, "What seemed to be adequate design, controls, and statistics."² He summed up 393 experimental comparisons on Television Vs. Classroom teaching including a considerable amount of published material. His report showed that of the 255 comparisons indicating no significant differences, 83 were significantly in favor of televised teaching and 55 significantly in favor of conventional teaching. Stickell in his widely quoted dissertation analyzed some 31 studies that made 250 comparisons between educational television and conventional instruction. He judged each comparison on the criteria of comparability of control and experimental subjects, assignment procedures, comparability of instruction, tenability of statistical assumptions, and adequacy of other controls. His findings showed, "Only 10 of these 250 comparisons were fully interpretable, and 21 did not meet two or more of the criteria and were uninterpretable. There were no significant


differences found in all 10 interpretable, and in 20 of the 23 partially interpretable comparisons."\(^1\)

In discussing the meaningful research and the resulting non-significant findings, Williams, Schramm, Skornia and Forsythe tend to respond to reasons other than instructional factors as the cause for such difference. Williams suggests, "That attitudes, ability and personality of the classroom teacher may be the most important missing variable."\(^2\) Schramm considers, "The Hawthorne effect (increased effort from the students due to being in the limelight as part of an experiment) or novelty-effect may have been important factors for non-significant different findings in closely allied fields." Skornia claims that,

Class-to-class and school differences are now so great as to invalidate much of the research at hand. Re-test should be considered before results are published. Studies will need to restrict themselves to measurement of factors that can be controlled and really measured so differences will be genuinely significant and generalizable.\(^4\)

---


\(^4\) Harry J. Skornia, "What We Know from New Media Research," NAEB Journal, XXV (March-April, 1966), 37.
Forsythe feels, "We may be asking the wrong questions and analyzing the wrong factors."¹ Williams concurs by stating "What appears to be virtual equality in effectiveness of televised and direct instruction may result from cancellation of different factors of effectiveness rather than from a similarity of such factors."²

By the end of the 1950's ITV trends were clearly established. It was no longer a question of whether to use television, but at Taylor comments, "The question was one of how to use television in Education."³ The initial exposure to television as a research tool developed a connection that the new medium could make a significant contribution to educational inquiry. One of the first well known training and diagnostic programs was micro-teaching in teacher education. Essentially constructed, but real teaching, the technique allows teachers to apply clearly defined teaching skills to carefully prepared lessons in a planned series of five to ten-minute encounters with a small group of real students, often with an opportunity to observe the results on video-


²Williams, loc. cit.

tape. Its distinction lies in the opportunity it provides teachers for immediate and individual diagnostic evaluation of teacher performance by colleagues, supervisors and participating students and for measuring progress in specific teaching techniques. Micro-teaching is used as an adjunct to either preservice teaching experience in the school or in-service programs of teacher improvement. This program was introduced at Stanford's teacher intern program in the Summer of 1963.¹

There was considerable support for the research potential of television recordings. In a study by Schueler and Gold at Hunter College New York, they reported that, "Television recordings are suitable records for developing objective measures of reporting teacher pupil behaviors."² Their research reported some success in developing a relatively objective observation schedule for describing teacher performance. It also indicated more must be done in order to develop a useful instrument for such observations. The greater potential use of the medium (video recordings), appeared to be in utilization with teachers already on the job rather than with student teachers.


Additional research has been conducted on the potential uses for television other than for the presentation of substantial parts of courses. This has been in the category of observation of demonstration teaching. Because of the logistics in taking increasingly large numbers of teacher trainees into classrooms to observe such situations, a number of studies have been directed toward comparisons of televised observations of teaching situations with actual classroom visits. With few exceptions, a non-significant difference was the primary finding.

Chabe in an experimental study at the State University of New York, Fredonia, was concerned with the effectiveness of observation of instruction via CCTV. During the course of the experiment eighteen junior-year elementary education students were enrolled in a social studies methods course during which they observed 25 fifth graders. Development of four objectives of social studies instruction were to be observed, (1) understandings such as concepts and generalizations; (2) skills, (3) attitudes and (4) appreciations. The primary objective of the study was to learn to what extent those objectives were being met in a selected fifth grade class. The college students were divided into two groups: (1) control (classroom observers) and (2) experimental (TV viewers). Chabe's major findings for the experiment indicated that televiewing was almost as effective as actual guided classroom observation. Guided televiewing
should be expanded and improved upon in teacher education with eventual, but not complete, replacement of actual classroom observation.¹

In almost identically titled studies by Mizer at the State University of Iowa, and Olivero from Stanford University came the idea of substituting video recordings for classroom observation. The purpose of Mizer's study was to use video tape recordings as a substitute for classroom observation when observations were impracticable or difficult to schedule. Olivero was concerned with a number of variables in substituting video tapes for classroom observations. These variables consisted of: (1) source of supervision, (2) type of feedback, and (3) condition of observation. Both studies employed the micro-teaching technique in their procedures. Mizer's study concluded that the greatest value of tape recordings seemed to be in helping students become familiar with teaching early in their preparation program. An additional finding was that most students accepted the use of tapes as a supplement to observation but not as a complete replacement.²


The Olivero study was less informative and simply stated that, "The results were conclusive to the degree that additional research is warranted. If the research continues to indicate positive improvement, the mechanical and administrative feasibility of the process may open new vistas for teacher education programs."\(^1\)

Observation by CCTV was used at the State University of New York, College of Education at Brookport. Rench, reported on the work there, and stated that six to ten children were observed at a time, and usually in the television studio only. The demonstrations lasted from twenty to thirty minutes, and covered essentially academic subject areas. The college students were provided with a mimeographed plan of the lesson before the observation, and provision was made for a post-observation discussion of the teaching methods observed and child responses heard.\(^2\)

Before the use of CCTV, the college had used the gymnasium floor and the auditorium to improvise studio classrooms. Rench continues:

> Based on observations of the college staff, and responses of students to periodic questionnaires, and opinions of campus-school teachers and children;


some of the advantages and disadvantages of CCTV\textsuperscript{1}. . . . The camera can bring closeup shots of children, their work, a picture, or a page of a book, thereby giving emphasis and clarity not possible in conventional classroom observations. When feasible, the college teacher can make comments and point out significant happenings at the moment they are taking place.\textsuperscript{2}

At San Jose State College, California, a three and one-half year research project was conducted to determine if the quantity of direct observation could be reduced by substituting CCTV observation. In the study, the control group made all of its observations by direct observation. The experimental groups, which were provided with CCTV observation, were also sent to the public school for varying amounts of their observation time—75 per cent, or 50 per cent, or only 25 per cent\textsuperscript{3} (see Figure 1).

<table>
<thead>
<tr>
<th>GROUP</th>
<th>OBSERVATION TIME</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental 1</td>
<td>TV 25%* 25% in-person</td>
</tr>
<tr>
<td>Experimental 2</td>
<td>TV 25% 50% in-person</td>
</tr>
<tr>
<td>Experimental 3</td>
<td>TV 25% 75% in-person</td>
</tr>
<tr>
<td>Control 1</td>
<td>100% in-person</td>
</tr>
<tr>
<td>Control 2</td>
<td>100% in-person</td>
</tr>
</tbody>
</table>

\textsuperscript{*}Refers to normal amount (50 hours) of in-person observation.

Fig. 1.—Observation time of experimental and control groups.

\textsuperscript{1}Ibid., p. 40.

\textsuperscript{2}Ibid., pp. 40-41.

\textsuperscript{3}William R. Roger, Television Utilization in the Observational Program for Teacher Education (San Jose, California: San Jose State College, 1962), pp. 1-58.
These experimental groups were then compared with two control groups who received no television observation but completed an established number of in-person observations. There was no difference among the five groups when each was compared using locally established rating scales. The findings indicated that observation of planned public school viewings via controlled television, plus decreased amounts of in-person observation was as effective as the total established amount of all five groups, including the two in-person observational control groups.

Rumford conducted an experimental study at the University of Colorado. One group of pupils was taught and observed directly and the other group of pupils was taught and observed by CCTV. The two groups of students were compared on the scores obtained from an objective test over their observations. When the students daily papers were used as a criterion, it was found that direct observation was less effective than the CCTV observation. When the objective tests made by the teacher were used, both methods of observation were found to be of equal effectiveness; however, when the objective test scores of the students with high ability were compared separately it was found that those students in the direct observational group did better.¹

It would seem that the results of the comparison of the two methods of observation would have been more meaningful had both groups of students been exposed to the same method of general course instruction rather than to expose one to direct instruction and the other to CCTV instruction.

In a previous study at San Jose State College, Clemens reported that, "Television observation conducted before in-person observation is significantly more effective than television observation only, in-person observation only, or television observation after in-person observation."\(^1\) The value of observations in a sequenced combination appears to be an important qualifying condition when comparing television and direct observations.

Observational treatment and approaches in using the media was a reoccurring theme in the following studies. Expanded enrollments at the University of Minnesota prompted some of the faculty to explore ways in which the University could better serve in its teacher education function. The College of Education (with the help of The Fund for the Advancement of Education and a grant from the National Defense Education Act, Title VII Project number 007), conducted studies in the utilization of direct observation, CCTV observation, kinescope recordings of observations, and instructional films.

\(^{1}\)Thomas D. Clemens, TV and Teacher Education (San Jose, California: San Jose State College, 1956), pp. 49-56.
Five Doctor of Philosophy dissertations were completed at the University of Minnesota in which various aspects of observation were evaluated and examined. Three studies by Gould, Thompson, and Abel were based on the following experiment. The college students enrolled in 'Introduction to Secondary School Teaching' were randomly assigned to three observational treatments: (1) direct observations, (2) CCTV observation, and (3) instructional films. The students made five observations by their assigned method and then made one observation by each of the other two methods. The CCTV observations were of speech and physics classes. The direct observations were made in the University High School and neighboring public schools by small groups of two and three students. The five instructional films used were from those produced by Indiana University and McGraw-Hill.¹

The structure of the study was such that all of the students were exposed to all forms of observations. All observation groups did not have an opportunity to view simultaneously the same lesson being taught. Since the five basic lessons observed by each group (public school, university high school, Indiana and McGraw-Hill films) were different, it does not seem possible to equate the observations.

The purpose of Gould's part of the study was to compare the three methods of observing on the basis of: (1) the students' evaluation of what each method contributed to the course instruction; (2) what the students considered to be educationally significant about each method; and (3) how the students responded on "standardized" classroom situations.

Gould's evaluations were based on summarized student reports, on significant teaching incidents, on a Film Test of Observation Skills, and on the Student Report on Observation Experience. He concluded, in contrast to Clemens, that the different methods of observation tended to supplement rather than replace each other; there was no real difference in what the three forms of observation tended to contribute to the instructional objectives; and, in total, the students tended to prefer the methods of observation in the following descending order—direct, CCTV, and film. The students generally preferred their dominate method of observation over the other two methods. Gould also found that there were no differences in perception activity as evidenced by the Film Test of Observation Skills. He qualified his judgment by noting that this may be because the measure was not sufficiently sensitive to the observation outcomes.\(^1\)

Thompson, in his part of the study, examined the possible relationships of professional attitudes and

\(^1\)Ibid., pp. 168-217.
interests to the three methods of observation. The Minnesota Teacher Attitude Inventory was used to measure attitudes toward instruction in the lecture and laboratory course; a Survey of Opinions was used to measure directiveness of instruction, and a Preferred Instructor Characteristic Scale was used to measure student-centered or subject-centered instruction. Thompson found that the different observation media did not affect the students' attitudes about children and that the students' attitudes were more closely related to differences in instructors than to differences in observation techniques.¹

Abel's part of the study was designed to measure the effect of the three methods of observation on the students' ability to apply their knowledge to classroom teaching situations and on the course subject matter outcomes. He evaluated the possible effects through the use of the Tests of Human Growth and Development by Horrocks and Troyers and through the use of course examinations. Abel stated that, with the exception of one standardized test, the instruments used in the study to evaluate subject matter outcomes were not sensitive to the three observational media. He

concluded that all three methods of observation could be used without loss of subject achievement.¹

Participating in another study at the University of Minnesota, Neale compared the effectiveness of direct observation, CCTV observation, and kinescope recordings of elementary school teaching. The use of several tests—of analysis of variance and covariance with the scores from the regular course examinations, the Minnesota Teacher Attitude Inventory, the Bowers Teacher Opinion Inventory, and a student rating scale for the course as a whole—resulted in no significant differences between the treatments. However, when the students rated their special lecture hour, those who made direct observation rated their experience significantly higher than those who made observations by CCTV and kinescope recordings, and those who observed by CCTV rated their experience significantly higher than those who made observations by kinescope recordings.² A test was not made specifically to measure what the students learned from their observations as Rumford had done in his study.


Adolphsen, in another study at the University of Minnesota, compared the effectiveness of various methods of observation—direct, CCTV, and kinescope, recorded as they had been used in an introductory course in secondary education. Half of the students assigned to the above three methods were also given supplementary direct observation. Many of the previous mentioned tests and inventories were used and a course examination was used to measure the students. An Observation Checklist and a Film Test of Teacher Perception were developed and also used as part of the study. The primary finding was that the nature and extent of the observation experience, rather than the media, were the significant elements in determining the perception of teacher behavior. It was also found that the student attitudes were more favorable toward direct observation; but the results indicated general acceptance and effectiveness of all methods of observation.¹

In summarizing the research findings of the preceding five studies, Keller found that, (1) for observational purposes there is no significant advantage of direct, CCTV and kinescopes of observation, and instructional films over

each other, and (2) combined observational methods are more desirable than individual ones.¹

In another related study using different conditions for observation Stroller observed eighteen elementary student teachers over a period of three semesters at Hunter College, New York.² Each of these student teachers was observed five times by the supervisors. The supervisors had three conditions of training: O, K, and OK. The O condition was the supervisor visiting the room in the usual fashion; no kinescope was made. In the K condition the supervisor observed the student teachers via closed-circuit and a kinescope was made at the same time. In the OK condition the supervisor visited the classroom, while the cameras and kinescope recorder also recorded the lesson. The results of the research on which supervisory condition was most productive of student teacher growth was inconclusive, but as a training tool the OK condition appeared to be the best of the three studied.

In another study by Brode and Ellis, forty-seven Northern Illinois University seniors trained in the use of


Flander's Interaction Analysis were used in observing classroom verbal behavior.\(^1\) The observations were made by classroom observation itself; by college CCTV and by audio only without watching the video portion. Again the non-significant difference was established as a result of the CCTV and Video Tapes when compared to in-person observation.

A more recent project on recording and evaluating progress of student teachers has been accomplished at the University of Utah. Harmer reports that,

> Ordinarily, evaluation of a student teacher follows a full teaching lesson. The teaching will be coded, and the evaluation will include all three individuals; the university professor, the cooperating teacher, and the student teacher.\(^2\)

This normal observing and conference technique was augmented by a second tool; namely, that of recording student teaching data on video tape. These video recorders were taken into the classroom, the student teacher's work was recorded and the video tape was later reviewed and analyzed. The primary intent of the Utah program has been to modify the student teacher's behavior. Another major purpose of the video

---

\(^1\)E. Leland Brode and Joseph R. Ellis, "The Effectiveness of Television Observation in Teacher Education," *NAEB Journal*, XXVI (March-April, 1967), Research Fact Sheet Series 1, No. 135, 1-3.

\(^2\)Earl W. Harmer, Jr., "Recording and Evaluating Progress of Student Teachers at the University of Utah," *Television and Related Media in Teacher Education* (Maryland: August, 1967), M-Step Monograph No. 1, pp. 51-52.
recordings has been to preserve student teaching data for subsequent research.

The role of the use of 'media' in terms of their vicarious application can be identified in the studies by Patrick;¹ Stoller, Lessor and Freedman.² Patrick's follow-up study showed greater long term gains are produced by media observations in direct expressions of teaching ability.³ This was true for both pre-service teaching and first year teaching performance.

A current article by Meier described the application of a combination of new media.⁴ This involved a six month training program for teachers of early childhood education scattered throughout the United States. Micro-teaching was used for making video tapes. These tapes were then used in a series of geographically dispersed satellites in Micro-training sessions. (For a discussion of Microtraining see


Meier.1) Observation and supervision were made by the satellite director who critiqued the trainee's work in accordance with previously established criteria. This was accomplished in some cases over a distance of 2000 miles.

In a non-instructional study using video tape recordings, a leading garment manufacturer has put video tape recordings into his training program. Perry reports:

Munsingwear has adopted video tape recordings with variable slow motion to modernize its training program and initiate new techniques in evaluating individual performances by observing sewing-machine operators.2

Munsingwear engineers objectively evaluated the sewing-machine operators on their productivity and rate of efficiency. This was accomplished by observing their techniques on video tapes using a television monitor. In investigating new uses for the recordings it was hoped that entire departments could be evaluated by video tapes rather than only observing employees on an individual basis.

Hughes reports the use of the video tape recorder in the area of athletics.

On February 3, 1968 the University of Washington gymnastic team was taped while competing against the Seattle Y.M.C.A. During a meet with the

---

1John H. Meier, "Rationale for and Application of Micro-Training to Improve Teaching," *Journal of Teacher Education*, XIX (Summer, 1968), 145.

University of Indiana on February 10, 1968 the University of Illinois recorded the performance of their team. These two tapes were sent to the president of the Mid-East Gymnastic Officials Association who had four neutral judges score the video taped routines in the same way a live gymnastics meet is judged.¹

Results of the scoring were not given in the article and it was mentioned that research on this problem was being conducted by a graduate student at the University of Washington at the present time. This article reflects the concept that observable traits in a specific activity apparently can be measured. The problems relative to the evaluation of observational research appear to be primarily related to the use of the 'hardware,' the method, and the design.

Summary

Because of the relative recency of television and related media as supplementing or implementing tools, communications research has been scarce in some aspects.

Instructional use of the media has dominated the literature. These instructional studies compared the performance of students instructed via television versus those instructed 'live' by the classroom teacher in the traditional manner. Inasmuch as this study was not primarily concerned with this type of research, and effort was made to limit

¹Eric Hughes, "Video Tape," *Journal of Health, Physical Education and Recreation*, XXXIX (May, 1968), 36.
these studies in the review of literature.

There has not been much research reported on the use
of the VTR as a supervisory tool. Those that have been done
have been observational studies directed toward comparisons
of televised observation of teaching situations against
actual direct classroom visits.

Presently, the reported research using television and
related media techniques, has tended to find no significant
differences between 'live' teaching situations versus the
taped teaching situations.

Heavy emphasis on instructional usage, as well as the
no significant difference syndrome characterizes the bulk of
the related literature.

While the present status of media research in the
preparation of teachers does not provide great illumination,
there does appear to be progress of a cumulative nature
to make this kind of research more comprehensive.
CHAPTER III

METHODS AND PROCEDURES

The six subjects who participated in this experiment were male physical education majors at The Ohio State University. They were actively engaged in their student teaching assignments during the Autumn quarter of 1968. The subjects were similar in their professional preparation having completed general and professional education courses, physical education requirements and pre-requisites for student teaching. The student teachers selected as subjects in this experiment were chosen from those student teachers under the supervisory direction of two university supervisors who also served as non-medium (live) raters during the experiment. This selection was accomplished randomly and was determined by the availability of the subjects, the raters, and the technician who taped all the sessions.

The subjects were observed by two rating groups. One group was called non-medium raters, and the other group was called medium raters. Each group consisted of three staff members from the Division of Men's Physical Education. The six raters were all experienced in undergraduate teacher preparation. Five of the raters held earned doctoral degrees and the sixth rater was a doctoral candidate.
The Non-Medium rating group consisted of:

Dr. Edward Coates, assistant professor and coordinator of student teaching, with additional duties as placement director for the Division of Men's Physical Education.

Dr. John Hendrix, associate professor and a university supervisor in the area of teacher education and preparation.

Mr. Arthur De Genaro, instructor with previous supervisory experience in teacher education and researcher in this study.

The Medium rating group was composed of:

Dr. Lewis A. Hess, Chairman of the Department of Men's Physical Education, whose specialization is in the area of teacher preparation.

Dr. Chalmer G. Hixson, Principal Supervisor of the undergraduate teacher education program for the division.

Dr. Charles L. Mand, Supervisor of the field experience program in recreation and active in teacher preparation programs. These two rating groups represented an aggregate total of 125 years of teaching experience or an average of 21 years per rater. These raters remained constant throughout the experiment.

The six taped sessions ranged from thirty-five to forty-five minutes in duration and represented a period of systematic observation covering two months. All observations were accomplished in sequence.
Each rater had a seven item check-list with a rating scale of score value ranging from 10 to 1, with 10 being the highest and 1 being the lowest. All major items of the check-list were considered to be of equal value (weight).

The seven areas rated were:

Area I. Personality and Personal Characteristics
Area II. Professional Qualities
Area III. Classroom Management
Area IV. Gymnasium Environment
Area V. Student/Teacher Relationship
Area VI. Future Success as a Teacher
Area VII. Non-Verbal Observation

Specific items under each of the major rating areas can be found in Appendix A.

The tapes were viewed only once by each of the medium raters. A tally sheet was developed for data collection (see Appendix B). This sheet contained cells where ratings were recorded and totaled. Scores of each individual rater for all seven items were totaled. Scores of raters A, B, and C, when summed, provided total composite scores of individual rating areas. The acceptance or rejection of the hypotheses in this study was bound to the agreement among the non-medium and medium raters and the composite scores between the two rating groups.

Prior to the actual taping, it was necessary to construct a rating scale check-list. Based on a review of
various check-list, inventories used at The Ohio State University, Kent State University, Temple University, State Teacher College at Brockport, New York, and Gage's standard reference, Handbook on Research and Teaching, a check-list was compiled and adopted.

An additional and extremely helpful source was a review by Wilson who conducted a survey of institutions of higher education in California to determine methods of evaluation and instruments used to evaluate teacher experience of student teaching. His study consisted of surveying thirty colleges and universities. This had been requested by the California Council on Teacher Education.\(^1\) Five of the headings included in each characteristic group from the top ten general characteristics and the top ten specific characteristics listed in his report were adopted for the seven major rating areas used in this study. As a result of Wilson's survey and based on the review of previously mentioned inventories from various institutions having teacher education programs, the check-list used in this study was given face and curricular validity. The number of areas to be included for an adequate check-list was the next step prior to actual taping. Symonds states, "While no rule can be given for the exact number of steps which should be

included on a rating scale, in most instances the use of seven units will yield optimal reliability.\(^1\)

Justification for the use of three raters was explained by Rugg. He maintains

> While experts differ in their recommendations concerning the number of persons whose judgments should be pooled, at least three independent ratings should be obtained in most instances in which human traits are being evaluated.\(^2\)

In a different report he comments further by saying, "If the three raters are so thoroughly acquainted with the person rated that they are competent to rate."\(^3\)

For the above reasons, as well as severe logistical considerations, three raters and a seven item rating scale for each group were chosen.

Singularly important in this study and pivotal to the experimental design was the hardware and apparatus involved with this research. The instrument used in conducting this research was a Sony Portable Video-Tape Recording Unit (see photograph Appendix C). The unit consisted of a CV-2000 Videocorder Deck and a solid state camera and normal lens. Also included were a wire microphone and a small television

\(^1\)Percival M. Symonds, Diagnosing Personality and Conduct (New York: Century Company, 1931), p. 79.

\(^2\)Harold O. Rugg, "Is the Rating of Human Character Possible," Journal of Educational Psychology, XIII (January, 1922), 81-93.

\(^3\)Harold O. Rugg, "Is the Rating of Human Character Possible," Journal of Educational Psychology, XII (November, 1921), 424-438.
monitor. The observations were recorded on five one-half inch V32 Helical Magnetic Tapes capable of recording one hour each. These tapes were provided on a loan basis from the American Association of Health, Physical Education and Recreation, Committee on the Utilization of Television in Physical Education. All other equipment described above was made available through the Division of Men's Physical Education Department, The Ohio State University.

Procedure

This experiment was conducted by formal systematic observation.

When predetermined units and time samples are used and when this method of observation is made systematically and objectively as possible these techniques gain in reliability and validity.¹

The format for collecting and compiling the data was conducted in two phases.

Phase One

Phase one consisted of the actual physical visits to the schools by the three non-medium raters and the technician who taped the sessions during these visits. The subjects were taped and concurrently evaluated by the non-medium raters. During these observation periods, the non-medium raters marked their check-lists where applicable and placed a

composite rating next to the major area heading. The data from all three non-medium raters were tallied on the data collection sheet after each taping. For identification purposes, the non-medium and medium raters were designated as rater A, rater B, and rater C. When all six subjects had been observed by the non-medium raters and the data from their check-lists had been transferred to the collection sheets, phase one was completed. This was the most time-consuming phase and needed the utmost coordination and liaison.

Phase Two

This phase began when the first tape was shown to members of the medium rating group. The identical procedure was followed for the recording of check-list scores and tallying of composite evaluations. These raters observed video taped recordings previously gathered during the non-medium phase of the same six subjects. Individual ratings were made by the medium raters without consultation. Although during some viewings more than one medium rater observed the same tape at the same time. After all medium raters viewed the entire six tapes in sequence, phase two was terminated.

Treatment of the Data

The data were collected and recorded for both groups of raters. The data obtained from the individual raters
within each group were analyzed for internal relationships and agreement by means of the Hotelling $T^2$. This was accomplished for all seven major areas of the check-list. Composite group scores of the non-medium and medium raters for all seven areas were administered the Parid $t$ for the purpose of determining agreement and symmetry between the two groups.
CHAPTER IV

ANALYSIS OF DATA

It was fundamental in the analysis of the data to provide a statistical treatment which would analyze raw data and evaluate differences among the individual non-medium and medium observational ratings. The Hotelling $T^2$ was a statistic which satisfied this requirement.\(^1\)

Additional statistical work was done which involved a composite analysis between non-medium and medium group rating scores. The significance of this kind of research was contingent upon statistical difference between comparative elements (variable 1, and variable 2). The variables were the total composite observational scores of the non-medium group and the medium group. The parid $t^2$ works well with the Hotelling $T^2$, and is an excellent comparative statistic; consequently, it was used for these reasons. The Hotelling $T^2$ dealt with differences among individual raters,


while the Parid measured agreement between the rating groups (variable 1 and 2).

The following design model is presented for additional clarification regarding the use of the Hotelling $T^2$:

<table>
<thead>
<tr>
<th>$R$ - Rater</th>
<th>$D$ - Difference</th>
<th>$R$-1</th>
<th>$R$-2</th>
<th>$R$-3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D_1$ = $D_2$ = 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(1) The difference between rater one ($R$-1) and rater two ($R$-2) is $D_1$.

(2) The difference between rater two ($R$-2) and rater three ($R$-3) is $D_2$.

(3) The mean difference of $D_1$ is equal to the mean difference of $D_2$ and also equal to zero.

This model was applied to both the non-medium and medium raters. The score values obtained by the Hotelling $T^2$ indicated the amount of difference from zero by both groups of raters.

A summary of the Hotelling $T^2$ applied to the evaluations made by the non-medium and medium raters is shown in Table 1.

The following discussion relates to the analysis of the Hotelling $T^2$ values for medium and non-medium raters. Acceptance of area one of the medium value from Table 1 was interpolated from the difference between $R$-1, and $R$-2, which was .773, and $R$-2, and $R$-3, which was 2.629. Both fell within the acceptance area and therefore, area one of
### TABLE 1
SUMMARY OF HOTELLING $T^2$ SCORES AMONG NON-MEDIUM AND MEDIUM RATERS WITH THE DIFFERENCE EQUAL TO ZERO

<table>
<thead>
<tr>
<th>Individual Rating Areas</th>
<th>df</th>
<th>Non-Medium</th>
<th>df</th>
<th>Medium</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2/4</td>
<td>.460</td>
<td>2/4</td>
<td>*</td>
</tr>
<tr>
<td>2</td>
<td>2/4</td>
<td>.237</td>
<td>2/4</td>
<td>2.544</td>
</tr>
<tr>
<td>3</td>
<td>2/4</td>
<td>7.898</td>
<td>2/4</td>
<td>.577</td>
</tr>
<tr>
<td>4</td>
<td>2/4</td>
<td>.400</td>
<td>2/4</td>
<td>.382</td>
</tr>
<tr>
<td>5</td>
<td>2/4</td>
<td>.669</td>
<td>2/4</td>
<td>.556</td>
</tr>
<tr>
<td>6</td>
<td>2/4</td>
<td>.331</td>
<td>2/4</td>
<td>.942</td>
</tr>
<tr>
<td>7</td>
<td>2/4</td>
<td>.148</td>
<td>2/4</td>
<td>2.529</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td>.385</td>
<td></td>
<td>1.349</td>
</tr>
</tbody>
</table>

**Notes:**

Critical values of the Hotelling $T^2$ for a one-tail test.

- Level of significance $0.025 = 10.65$
- Level of significance $0.05 = 6.94$

**df**: $N-1$ Three raters: $3-1 = 2$ Six subjects: $6-2 = 4$

$df = 2/4$.

*The Hotelling $T^2$ could not be computed due to the singularity of the covariance matrix (the divisor was zero).*
medium ratings was assumed to fall into the acceptance area also (see Figure 2)

<table>
<thead>
<tr>
<th>Critical Values</th>
<th>Acceptance Area</th>
<th>Rejection Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.65 = .025</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.94 = .05</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 2.—Hotelling $T^2$ acceptance and rejection areas.

The third area rated by the non-medium group (classroom management) fell in the .05 rejection area with a value of 7.898. This resulted from non agreement between non-medium rater R-2, and rater R-3, with a value between them of 4.392. When compared with R-1, and R-2, which was 1.913: and when the Hotelling $T^2$ was applied, the value did not fall into the .05 acceptance area for the non-medium raters for item three. However, it did fall within the .025 level.

Attention must be focused on the non-medium value placed on the seventh rating area (non-verbal observation). The agreement of .148 was well into the acceptance area for this highly subjective rating area. This indicated both agreement and symmetry among the non-medium raters for this area.

The total values of the Hotelling $T^2$ for the non-medium raters (.385), and the medium raters (1.349), were both well within the acceptance area.
Null Hypotheses (one and Two)

1. There will be no agreement in the evaluations of the subjects by the three raters observing by direct observation (non-medium ratings).

The hypothesis was rejected on the basis of acceptance for all rating areas. Number three can be accepted at (p < .025 level). The total non-medium value of .385 tends also to support a rejection of the null hypothesis.

Based on these empirical findings, as well as other related statistical support, the alternative hypothesis is accepted. This alternate hypothesis states: There will be agreement in the evaluations of the subjects by the three raters observing by direct observation (non-medium ratings). A summary is presented in Table 1.

2. There will be no agreement in the evaluations of the subjects by three different raters observing by means of video tapes (medium ratings).

This second hypothesis was also rejected. The basis for the rejection was agreement by all medium ratings for the values of all seven rating areas as well as the total value rated by the medium group.

Again based on the findings, the alternative hypothesis was accepted. It states that: There will be agreement in the evaluations of the subjects by three different raters observing by means of video tapes (medium ratings). Refer to Table 1 for results.
The next step was to analyze the difference between the composite non-medium scores and the composite medium scores. The acceptance of the major hypothesis was contingent upon the computed statistics (Parid t), in conjunction with a one-tail test of significance. (The Hotelling T² previously used for agreement among raters extends the Parid t for use in determining symmetry between the rating groups.)

A one-tail test was used for testing the hypothesis that the mean difference was zero or negative. The one-tail test is appropriate for power reasons. The power of the test is increased in that a one-tail test is more sensitive to differences when the difference is genuine.¹

Acceptance and rejection areas for the Parid t used in a one-tail test can be seen in Figure 3.

Critical values

<table>
<thead>
<tr>
<th></th>
<th>Acceptance</th>
<th>Rejection</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.774</td>
<td>.01 Area</td>
<td></td>
</tr>
<tr>
<td>3.163</td>
<td>.05</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 3.—Parid t acceptance and rejection areas.

A distinctive feature of this study relates to the confidence levels. Confidence levels of .05 and .01 were

reported for the Parid. This decision was based on several factors:

1. The sample and size of the N......"If it is not possible to demonstrate that a difference exist, it is probably not large enough to be of practical significance and/or the sample size was not sufficiently large enough."\(^1\)

2. The lack of major related studies which would provide agreement and establish precedent for selection of specific confidence intervals.

3. The hardware and its use by the raters were factors for consideration in reporting multiple confidence intervals.

In using the new media some learning may have taken place on the part of the medium raters; such as what to look for on the tapes, technical areas of adjusting the monitor for sound level, clarity of picture, and in general how to observe by means of video tape recordings.

4. Composition of the rating scale was also an important factor in the choosing of a confidence interval. A greater degree of latitude in the selection of a confidence interval was necessary, because both subjective and objective areas were included to be rated. Also specific elements within each of the general rating areas must be clearly identifiable to the raters prior to the setting of definite confidence limits.

\(^{1}\text{David V. Huntsberger, Elements of Statistical Inference (Boston: Allyn and Bacon, Inc., 1963), p. 143.}\)
A summary of the Parid applied to these composite group scores is shown in Table 2. Scores included in this table reflect composite standard deviation and mean differences for the Parid as well as standard deviation and mean scores for non-medium and medium groups.

The analysis indicated that areas one through six as well as the total of the Parid fell into the acceptance area (based on .05 and .01 levels of significance). The seventh area (non-verbal observation), a highly subjective and difficult area to rate, fell in the rejection area with a value of 8.800. This was due to the wide non-agreement between non-medium ratings (.148) and medium ratings (2.529) which when the Parid was applied fell into the rejection area.

Throughout the experiment a yardstick developed regarding the medium ratings. There was a consistently higher degree of symmetry in the medium (taped observations), as reflected by the mean differences and mean scores for the medium group. This measure of consistency is evidenced in all seven rating areas.

Based on the data from this table, the following observations can be made: (1) rating areas one through six (Personality and Personal Characteristics, Professional Qualities, Classroom Management, Gymnasium Environment, Student/Teacher Relationship, and Future Success as a Teacher), all fall within the acceptance regions established
TABLE 2  
SUMMARY OF PARID t DIFFERENCE BETWEEN NON-MEDIUM AND MEDIUM GROUPS

<table>
<thead>
<tr>
<th>Rating Areas</th>
<th>df</th>
<th>Parid t</th>
<th>Mean Non-medium</th>
<th>Mean Medium</th>
<th>Mean Difference</th>
<th>S.D. Non-medium</th>
<th>S.D. Medium</th>
<th>S.D. Composite</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>3.616</td>
<td>17.417</td>
<td>20.583</td>
<td>2.167</td>
<td>3.541</td>
<td>3.774</td>
<td>2.145</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>2.108</td>
<td>18.317</td>
<td>20.167</td>
<td>1.850</td>
<td>3.371</td>
<td>3.656</td>
<td>2.149</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>2.873</td>
<td>18.217</td>
<td>20.917</td>
<td>2.700</td>
<td>2.846</td>
<td>3.800</td>
<td>2.302</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>1.124</td>
<td>19.050</td>
<td>20.167</td>
<td>1.117</td>
<td>5.458</td>
<td>4.203</td>
<td>2.433</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>2.108</td>
<td>18.967</td>
<td>20.983</td>
<td>1.867</td>
<td>3.951</td>
<td>3.940</td>
<td>2.170</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>8.800</td>
<td>18.117</td>
<td>21.333</td>
<td>3.217</td>
<td>3.891</td>
<td>3.342</td>
<td>.283</td>
</tr>
</tbody>
</table>

Notes:

Critical values of the Parid t for one-tail test of significance
df: N-1 Six observations
6-1 = 5 df.

Level of significance .01 = 4.774
Level of significance .05 = 3.163
by the confidence intervals. (2) Individual Parid $t$'s as well as the total Parid $t$ score values, fall within the acceptance region established by the confidence intervals. (3) Area seven (non-Verbal Observation), fell into the rejection area as established by the confidence levels for the Parid $t$ design. However, both the non-medium and medium ratings when taken individually and evaluated by the Hotelling $T^2$ design, fell well within the acceptance areas of that design. This statistic suggests internal agreement, but a lack of symmetry when the ratings were compared by the Parid $t$.

**Major Null Hypothesis**

There will be no agreement between composite evaluations of subjects observed and rated by direct physical observation (non-medium ratings), and the composite evaluations of those same subjects when observed by means of video tapes (medium ratings).

Based on the statistical evidence and findings of the data, the null hypothesis has been rejected. Consequently, the alternative hypothesis is accepted. This hypothesis states: There will be agreement between composite evaluation of subjects observed and rated by direct physical observation (non-medium ratings), and the composite evaluations of those same subjects when observed and rated by means of video tapes (medium ratings).
CHAPTER V

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

The purpose of this study was to determine, by formal systematic observation, the capability of the video tape recorder as an evaluative instrument in supervision. The problem was to determine, by experimental method, the agreement and symmetry between non-medium and medium composite observational scores. The need for this type of study was to improve the effectiveness of university staff personnel in the evaluation and supervision of student teachers in physical education.

The need for greater staff utilization and economy of time involved in the physical visitations by the staff member were primary considerations of the study. The administrative and educational implications were to increase the frequency of observations while decreasing the physical visits by the university supervisor.

The unique quality of this study is based on the sparseness of research in using the Video Tape Recorder as an evaluative tool rather than as a means of instruction (see Chapter II--instructional category of 'Relative Effectiveness Studies').
The subjects were six physical education student teachers from the Division of Men's Physical Education at The Ohio State University. They were observed by two rating groups consisting of three raters each. One group was called non-medium raters, and the other group was called medium raters. Each group consisted of three staff members from the Division of Men's Physical Education. The six raters were all experienced in undergraduate teacher preparation. The non-medium raters observed by means of physical or direct observations. The medium group raters observed by viewing the subjects by means of video tape.

Each rater within the rating groups used an identical check-list for scoring his evaluations. Prior to the actual taping, it was necessary to construct a rating scale check-list. Based on a review of various inventories used at a number of institutions engaged in the preparation of teachers, a check-list was adopted and given face and curricular validity. An additional and extremely helpful source was a review by Wilson who conducted a survey of institutions of higher education in California to determine methods of evaluation and instruments used to evaluate teacher experience of student teaching.

Six teaching sessions were taped ranging from thirty-five to forty-five minutes in length. The tapes were produced by a Sony CV 2000 Video Tape Recording Unit using one-half inch V 32 Helical Magnetic Tapes. Equipment was made
available through the Division of Men's Physical Education Department, The Ohio State University, and the Committee on the Utilization of Television in Physical Education of the American Association for Health Physical Education and Recreation.

When the data were collected, the totals were recorded on a tally sheet (see Appendix B). The two statistical procedures applied to the data were the Hotelling $T^2$ and a Parid $t$. First, the Hotelling $T^2$ was applied to the data in order to analyze the non-medium and medium group scores (raw scores) for internal relationship and agreement. This was followed by a Parid $t$ for the purpose of determining agreement and symmetry between the composite group scores of the non-medium and medium rating groups. (Complete results concerning these two statistical treatments can be found in Tables 1 and 2.)

A one-tail test was used for testing the hypotheses that the mean difference was zero or negative. The one-tail test is appropriate for power reasons. The power of the test is increased in that a one-tail test is more sensitive to differences when the difference is genuine.

Acceptance and rejection areas were established on the baseline of the one-tail tests. In this study, because of reasons stated previously, multiple acceptance and rejection areas were used. The acceptance area refers to those scores which fall within the critical values established for the Hotelling $T^2$ and the Parid $t$ used in this
study. The rejection area is that region on the baseline which excludes a significant score at the appropriate confidence limit set by the design used.

Several confidence intervals were reported for the Hotelling $T^2$ and Parid $t$. This technique was used for ease of inspection of the tables for the interested reader. Rather than structure the statistics with a single confidence interval, the multiple technique was used.

Major Conclusions and Recommendations

Major conclusions and recommendations of this study based on the statistical interpretation of the data indicated the following:

1. There appears to be sufficient agreement between the scores of the non-medium and medium groups, as reflected by the Parid $t$ scores, to indicate implementing the Video Tape Recorder into programs of teacher preparation for supervisory functions.

2. The use of VTR is an important facet in the total evaluation of physical education student teachers. This would replace other forms of observation entirely, but would supplement them in a much more effective manner.

3. Such items as voice and physical appearance can be observed with varying degrees of agreement by medium ratings. This appeared to be true even in the areas where items rated were subjective in nature. (See Table 1, Chapter IV—Rating area seven: Non-Verbal observation.)
4. Based on the statistical rejection of the three hypotheses and acceptance of the alternate hypotheses, support is given to the concept of the VTR as a means of evaluating and supervising student teachers of physical education.

5. Further media research in the use of Video Tape Recordings for supervision and evaluation of student teachers in physical education is warranted and needed.

6. Based on the technical conclusions (that follow), it is recommended that an interdisciplinary approach be pursued for the taping of student teachers of physical education. This might take the form of a required or elective course in field experience for a student specializing in the communications field. He would be responsible for technical replication (by means of VTR) of the physical education student teaching experience.

Technical Conclusions

The ideas presented in this section of the paper evolved from the actual 'doing' of the experiment rather than from the statistical findings. These technical conclusions are:

1. A technician should be employed to tape the observations because of his expertise in the use of the hardware.

2. A technical (software) conclusion is that for better quality and reproduction of the tapes a zoom lens along with a cordless or remote microphone system is recommended.
3. Prior to incorporation of a program of taping student teaching observations, all parties involved, i.e., students, cooperating teachers and their schools, university supervisors, and other staff members who are in the teacher preparation area should be made completely aware of what the program is attempting to accomplish.

4. The problem of logistics must be considered before implementing a supervisory program using Video Tape Recordings. Formal systematic observation requires adequate personnel for taping and observing, readily available equipment and materials, and coordination among the schools where observations are to be conducted.

5. The rating scale and check-list must reflect specific areas for evaluation so as to provide the observer with meaningful evaluation guidelines concerning the student teacher under observation.
APPENDIX A

CHECK-LIST AND RATING SCALE
Check List

Directions:

Rating scale will consist of values 10 through 1, with 10 being the highest and 1 being the lowest; 5 is to be considered average or mid-range. There can be a breakdown within each value to indicate finer markings, i.e., 8.5 or 6.9. When marking, it is suggested that the rater think in terms of percentages so as to derive accurate and meaningful markings. A list of items will fall beneath each major check-list division. These will help the rater in making his final marking. The rater is not to give a numerical rating on the check-list, but only to check off the items if observed. Only the major heading areas are to be given numerical rating.

<table>
<thead>
<tr>
<th>Check-List</th>
<th>Composite Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>AREA I. Personality and Personal Characteristics:</td>
<td></td>
</tr>
<tr>
<td>___ a) Quality of Voice</td>
<td></td>
</tr>
<tr>
<td>___ b) Language and use of English speech</td>
<td></td>
</tr>
<tr>
<td>___ c) Neatness and general appearance and posture</td>
<td></td>
</tr>
<tr>
<td>___ d) Emotional control</td>
<td></td>
</tr>
<tr>
<td>___ e) Preparation and presentation of subject</td>
<td></td>
</tr>
<tr>
<td>___ f) Sincerity and empathy</td>
<td></td>
</tr>
<tr>
<td>AREA II. Professional Qualities:</td>
<td></td>
</tr>
<tr>
<td>___ a) Ability as a teacher</td>
<td></td>
</tr>
<tr>
<td>___ b) Ability to demonstrate</td>
<td></td>
</tr>
<tr>
<td>___ c) Knowledge of subject area, teaching techniques</td>
<td></td>
</tr>
<tr>
<td>___ d) Motivation and leadership ability</td>
<td></td>
</tr>
<tr>
<td>___ e) Awareness of individual</td>
<td></td>
</tr>
<tr>
<td>___ f) Awareness of laws of learning and growth and development patterns</td>
<td></td>
</tr>
<tr>
<td>AREA III. Classroom Management:</td>
<td></td>
</tr>
<tr>
<td>___ a) Organization of class and equipment--to include audio/visual aids</td>
<td></td>
</tr>
<tr>
<td>___ b) Discipline</td>
<td></td>
</tr>
<tr>
<td>___ c) Awareness of time space factors</td>
<td></td>
</tr>
<tr>
<td>___ d) Flexibility and adaptability</td>
<td></td>
</tr>
<tr>
<td>___ e) Use of teaching assistants, squad leaders</td>
<td></td>
</tr>
</tbody>
</table>
Check-List

AREA IV. Gymnasium Environment:

a) General overall facilities
b) Safety hazards within the teaching area
c) Use of proper safety measures, i.e., mats, belts, chalk, spotters
d) Condition of equipment and items of wearing apparel

AREA V: Student/Teacher Relationship

a) Empathy towards children
b) Sense of humor
c) Cooperation
d) Rapport with class (friendly, fair, and firm)
e) Class responsiveness
f) Motivation and leadership

AREA VI: Future Success as a Teacher:

a) Potential as a teacher
b) Professional traits, interest, growth
c) Dedication and Professionalism
d) Relations with cooperating teacher and university supervisor

AREA VII: Non-verbal Observation:

a) Effort (tries hard)
b) Attitude
c) Covert behavior
d) Role playing as a student teacher
<table>
<thead>
<tr>
<th>Subject No.</th>
<th></th>
<th></th>
<th></th>
<th>TOTAL COMPOSITE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MEDIUM RATER</td>
<td>RATER</td>
<td>RATER</td>
<td>RATER</td>
</tr>
<tr>
<td>Check List</td>
<td>RATING AREA</td>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX C

PHOTOGRAPH OF SONY PORTABLE RECORDING UNIT
BIBLIOGRAPHY
BIBLIOGRAPHY

Books


Rogers, William R. Television Utilization in the Observational Program for Teacher Education. San Jose, California: San Jose State College, 1962.


Articles


Skornia, Harry J. "What We Know from New Media Research." NAEB Journal, XXV (March-April, 1966), 37.
Stoller, Nathan, Lessor, Gerald S., and Freedman, Philip I. "A Comparison of Methods of Observation in Pre-
Service Teacher Training." AV Communication Review, II (Summer, 1964), 177-197.

Williams, Catherine M. "Reexamination of No Significant
Differences that ITV Studies Report." AV Communication

Wilson, Donald E. "A Survey of Evaluation Instruments and
Procedures Used in California Teacher Training Insti-
tutions During Directed Teaching Experience." Journal
of Educational Research, XLVIII (May, 1955), 649.

Reports

Allen, Dwight W. Micro-Teaching: A Description. Report of
Stanford Teacher Education Program (STEP), Stanford: Stanford University, Stanford, California, 1967.

Clemens, Thomas D. TV and Teacher Education. San Jose,
California: San Jose State College, 1956.

Chu, Godwin C., and Schramm, Wilbur. Learning from Televis-
ion: What the Research Says. A Report by the Insti-
tute or Communication Research, Stanford: Stanford
University, Stanford, California, 1967.

Harmer, Earl W. Jr. Television and Related Media in Teacher
Education. "Recording and Evaluating Progress of
Student Teachers at the University of Utah," Maryland:
1967, M-Step Monograph No. 1.

Keller, Robert J. Closed Circuit Television in Teacher Edu-
cation, A Graphic Report, 1961. Minneapolis, Minnesota:
University of Minnesota, College of Education, Uni-
versity High School, 1961.

Patrick, Robert B. The Measurement of the Effectiveness of
the Documentary Sound-Film as a Supplement in Teaching
of Methods to College Students Being Prepared to
Teach in the Secondary Schools. University Park,

Patrick, Robert B., and Davison, Hugh M. The Effectiveness
of the Documentary Sound-Film as a Supplement in Secon-
dary School Teacher Education: A Follow-up of First
Year Teacher Performance. University Park, Pennsyl-


Newspaper

Columbus Dispatch Supplement. September 29, 1968.

Unpublished Material


