8-2009

The Effects of Training on Preservice Teachers using Computerized Systematic Observation

Stephen T. Marriott
The College at Brockport, SMar1120@yahoo.com

The Effects of Training on Preservice Teachers using Computerized Systematic Observation

by

Stephen T. Marriott

A thesis submitted to the Department of Kinesiology, Sport Science, and Physical Education College at Brockport State University of New York in partial fulfillment of the requirements for the degree of Masters of Science in Education (Physical Education)

August 2009

The Effects of Training on Preservice Teachers using Computerized

Systematic Observation

by: Stephen T. Marriott

APPROVED BY:

Alisa R. James — 11-11-09

Thesis Chairperson: Dr. Alisa R. James — Date

Danny Too — 11/11/09

Committee Member: Dr. Danny Too — Date

Douglas H Collier — 11-11-09

Committee Member: Dr. Douglas H. Collier — Date

Susan C. Petersen — 11/11/09

Chair, Department Kinesiology, Sport Science, and Physical Education — Date

Dr. Susan Petersen

Table of Contents

List of Figures

List of Appendices

Abstract

Although systematic observation software systems are being used in teacher preparation programs, research investigating the type and amount of training pre-service educators need to use this software to code teaching behaviors is lacking. The purpose of this study was to investigate the differences in three training protocols for preservice physical educators using the Behavioral Evaluation Strategies and Taxonomy software system for the first time. Participants were 31 pre-service physical education teacher education students enrolled in a *methods of elementary physical education instruction course* at a midsized college located in western New York State. Data were collected using a function from the BEST software system that automatically charted frequencies of each behavior recorded by the user. Data were analyzed using SPSS (version 17.0). Descriptive statistics were obtained and an ANOVA was used to determine whether there were differences (and level of significance) between four different training group means. Each experimental group was compared to the control group using Dunnett post-hoc tests. An unpaired t-test (two-tailed) was used to determine whether there were differences (and level of significance) between the participants who received a training video and the participants who did not. Results of ANOVA determined differences to be significant at $p = 0.060$ between the four groups. Dunnett post-hoc tests determined significance levels for the following comparisons between the Control Group (CG) and Training Protocol 1 ($p = 0.284$), CG and Training Protocol 2 ($p = 0.041$), and CG and Training Protocol 3 ($p = 0.075$). Results of the unpaired t-test (two-tailed) indicated participants viewing the training video increased their ability to identify a greater amount of feedback at $p = 0.025$. The results of this study suggest using video training

techniques to train preservice physical education teachers to use systematic observation software.

Chapter 1

Introduction

While research measuring teacher effectiveness has been conducted extensively in physical education using a variety of systematic observation instruments, there has been little research to determine the amount and type of training necessary for preservice physical educators to accurately code teaching behaviors using computerized systematic observation software. Furthermore, the research that has been completed has not examined physical education preservice teachers' ability to systematically code effective teaching behaviors using computerized systematic analysis software such as the Behavioral Evaluation Strategies and Taxonomy (BEST). The BEST software system has been used previously in teacher education programs as an appropriate method to evaluate teacher education students coding teaching behaviors (Heath, Coleman, Lensegrav, & Fallon, 2006; James, 2008).

In teacher education programs, it is not practical for students to reach a reliability standard that may take forty or more hours of training to attain (Deng Keating, 1999). The literature is unclear with regard to how much and what type of training is necessary for preservice teachers to successfully use computerized systematic observation software.

Systematic Observation as a Tool

Systematic observation has allowed individuals to examine themselves and investigate effective teaching behaviors with regard to their own teaching (Behets, 1993, Maeda, 2001, Kahng & Iwata, 1998). Teachers, coaches and administrators have used systematic observation tools in educational settings for the purpose of supervising inservice teachers, training preservice educators, as well as modifying individuals' teaching strategies. Many of these tools, proven to be valid and reliable, have been outlined by Darst, Zakrajsek, & Mancini (1989). When used properly, systematic observation can enhance teaching effectiveness.

Techniques / Strategies in Systematic Observation

Systematic observation allows educators to identify areas of strengths and weaknesses in regard to teacher and student behavior and to develop a self-evaluative system to assist them in modifying their teaching patterns. Areas that teachers may evaluate using systematic observation tools include: practice time, instruction time, class management, response latency, student performance, instructional feedback, student contacts, and active supervision (Darst & Pangrazi, 2005).

Traditional paper and pencil methods of systematic observation have included both qualitative and quantitative methods of gathering data. Qualitative methods include methods such as 'eyeballing' and anecdotal recording. 'Eyeballing' refers to an outside observer examining educational variables (i.e. teachers, students) without any written record of what was seen. For example, a teacher may have a peer observe how the instructor interacts with a particular student in order to provide insight and provide suggestions based on the observations. Anecdotal recording is similar to eyeballing except that the observer records his/her observations. Anecdotal recording involves a written record of progress based on milestones in teacher development (American Association of School Administrators, 1992). If done accurately, anecdotal recording procedures can provide a true and unbiased account of what is occurring in the gymnasium.

Quantitative methods of systematic observation such as rating scales and checklists provide a numerical value associated with an observation. Rating scales and checklists can be developed to obtain data about the frequency, duration, intensity, and/or latency of a certain occurrence of a behavior. Rating scales and checklists can be effective when examining such specific teaching behaviors as: (a) the number of demonstrations in a lesson, (b) the amount of feedback provided, or (c) the number of times a teacher uses students' names.

Both qualitative and quantitative methods can be unreliable and become more so when there are more variables to be observed. To enhance reliability, observers have begun to use computerized systematic observation software systems. Furthermore, the use of technology has allowed for the evolution of computerized versions of systematic observation tools that have been shown to be a viable alternative for data collection and analysis (Deng Keating, 1999).

The Six Critical Steps to Systematic Observation

Systematic observation techniques follow a certain process that is far more complex than watching lessons and collecting data on a few selected behaviors and events. Darst & Pangrazi (2005) suggested a process of systematic observation that is comprised of six critical steps. First, an individual must decide *what to observe*. A specific focus or goal is identified, which is based on the values of who is doing the observing. Deciding on what is to be observed needs to be a collaborative effort involving teachers, program leaders, and researchers in the area of systematic observation (Darst et al., 1989). For example, teacher-training programs may have preservice teachers looking at specific management techniques or the total amount of management time during a lesson. Staff development programs may have teachers systematically observe their own teaching and record the amount of feedback given to male students and compare that data to the amount of feedback given to female students. Other examples include the amount of feedback provided to students of various skill levels or the amount of questions used in a lesson.

Secondly, *specific definitions* need to be developed with regard to each behavior or event observed. Precise, operational definitions assist the observer in delineating similar behaviors or occurrences and can minimize disagreement leading to poor reliability between and/or within observers (Darst et al., 1989). Hawkins, 1982, has outlined the components for a complete definition of the behavior or event to be observed. Hawkins recommended that each definitions should have the following components: (a) descriptive name, (b) general description, (c) elaboration that describes the critical parts of the behavior, (d) typical examples of the behavior, and (e) questionable / borderline or difficult examples of both occurrences and non-occurrences of the behavior.

The third step in systematic observation is *selecting the most appropriate observation tool* and determining if there is an existing observation tool that fits the need of the observer. Once the definitions are in place the behavior has to be characterized one or more of the following: (a) frequency, (b) intensity, (c) duration, (d) latency, (e) endurance, or (f) accuracy (Bailey and Wolery, 1989). Frequency, or rate, refers to how often a behavior occurs. Some behaviors may occur multiple times, others may occur infrequently, while others may form patterns preceding a specific behavior (Darst et al., 1989). Intensity refers to the amount of force with which the behavior occurs.

This is also referred to as amplitude. Duration refers to the length of time a given behavior lasts. Latency refers to how long it takes a child to initiate a behavior once it has occurred. Endurance refers to the length of time a given behavior can be repeatedly performed. Accuracy refers to the extent to which a child's behavior conforms to the defined topography of a given behavior (Bailey and Wolery, 1989).

The fourth step involves establishing *observer reliability*. Observer reliability is an important aspect of systematic observation and needs to be established when coding videotape. By definition, reliability refers to the capacity of the instrument to yield consistent scores or results during multiple trials (Johnston & Pennypacker, 1980). Observer reliability can occur between different observers (interobserver reliability) or within one's own observations (intraobserver reliability). High levels of reliability can be more likely accomplished by using a sound training protocol. It is recommended that both intra and interobserver reliability reach an agreement of 80% (Darst et al., 1989).

The fifth step concerns an *awareness of the Hawthorne effect*. The Hawthorne effect occurs when the presence of observers, video camcorders, or audio recording devices in the gymnasium influence student or teacher behavior, thus inappropriately influencing the results. It has been suggested that students will become accustomed to the observers and devices being used and participant behavior will return to its regular pattern (Thomas, Nelson, & Silverman 2005). In order to combat the Hawthorne effect, the observer and equipment used should be as inconspicuous as possible and even out of sight using unobtrusive research techniques such as glass mirrors. In addition, if the individuals being observed are provided with prolonged exposure to the equipment and observers, they may become accustomed to the observation. Thus, consistency in the placement of the equipment needs to be considered when repeated observations are being made.

The final step concerns *summarizing and interpreting the data*. Data should be displayed in a manner that provides feedback effectively to the instructor. Data can be displayed in several different ways. Examples include a pie chart, frequency count bar graphs, and scalable time plots. These approaches demonstrate frequency and duration distribution in a graphic manner.

Computer Technology

Over the past two decades, advances in computer software and the use of mobile personal computers have led to the development of computerized systematic observation systems. Systematic observation techniques utilizing computer software and hardware have enhanced systematic observation by improving the reliability and accuracy of recording. In addition, computer approaches have improved the efficiency of data calculation and graphing for systematic observation (Donat, 1991; Eiler, Nelson, Jensen, & Johnson, 1989).

Advances in technology also have led to more user-friendly computerized systematic observation programs. For example, the Behavior Observer System (BOS) uses handheld computers whereby behavioral data was entered by touching buttons located on the screen. This program allowed users to easily and accurately record behaviors such as: (a) response frequency, (b) duration, (c) intervals, (d) time samples, (e) latency, (f) interresponse time, and (g) discrete trials. A second program, the Direct Observation Data System (DODS) has the capability to capture frequency, duration, interval, time sample, latency, and antecedent-behavior-consequences data (Kahng &

Iwata, 1998). In addition, the BOS software program includes a reliability statistical program to determine inter and intraobserver reliability while the DODS software program does not have that capability. Although the BOS and DODS are both effective software systems, a third software program, the Behavioral Evaluation Strategies and Taxonomy (BEST) combines many features from both programs and is extremely user-friendly.

This BEST software system is split up into two programs; one to collect data and one to analyze data. The first part of the program, the BEST Collection, allows the user to precisely define the variables of interest. For example, the BEST Collection software allows the user to conduct duration recording for the following variables: (a) instruction, (b) management, (c) activity, and (d) waiting time. Also, it allows the user to code the number of times each of the following occurs: (a) use of student names, (b) specific congruent feedback, (c) general feedback, (d) corrective feedback (e) positive behavior feedback, (f) negative behavior feedback, (g) demonstrations, and (h) the amount of times the instructor asks questions of the student(s). The second component of BEST is used to view data. This component of the software automatically tallies data collected and can generate multiple graphs and charts to view the frequency of each behavior coded.

Problem Statement

It is unknown how much or what type of training is needed for preservice teachers to code data using computerized systematic observation techniques such as Behavioral Evaluation Strategies and Taxonomy software system. While preservice teachers are enrolled in methods of instruction courses, they are learning what systematic observation is, how to do it, as well as specific variables which they may code using systematic observation tools. Although learning about systematic observation is a common aspect of curricula for preservice teachers, an effective method to train preservice teachers to use systematic observation tools has not been identified in the literature.

Purpose of the Study

The purpose of this study was to investigate the differences amongst training protocols used to train preservice physical education teachers as first time users of the BEST software system.

Research Questions

Question #1: What are the differences between the control group and the experimental groups in coding the video (CV)?

Question #2: What are the differences between participants who viewed the training video and participants who did not?

Question #3: Are physical education methods classes providing enough instruction about effective teaching behaviors and training in systematic observation for preservice educators to successfully systematically observe and code teaching behaviors?

Assumptions

In order for the training protocol to be successful, it was assumed that the participants gave their full attention while instruction took place. It was assumed that participants had no prior experience with systematic observation software programs. It was also assumed that participants were able to perform basic computing functions.

All participants were enrolled in two separate sections of *methods of elementary physical education instruction courses* that were taught by two different professors. It was assumed that, although their methodologies may differ, the professors taught the same curriculum with regard to systematic observation.

Delimitations

Participants were recruited primarily from a midsized college located in western New York State. All participants were students enrolled in a *methods of elementary physical education instruction course*.

Limitations

The primary limitation in this study is the coding video (CV). The coding video in the current study used two preservice teacher education students (one male, one female) with minimal teaching experience. The students who participated in the lesson were also preservice teacher education students. An undergraduate student commanded a single video camera and this student was responsible for videotaping both teachers. The videotape often switched between the two instructors as well as the students in the class. In addition, audio was picked up using the microphone on the video camera. Loud noises occurring during the lesson (hockey sticks banging on the ground, students talking close to the camera, music during the lesson) sometimes made the instructors difficult to hear.

Significance of Study

Little is known about the types and amount of training preservice teachers need to successfully use systematic observation techniques. There are inconsistencies in the research literature regarding the amount of training needed for first-time users of systematic observation instruments (Behets, 1993, Deng Keating, 1999, McKenzie, Sallis, & Nadar, 1991). It has been determined that the pedagogical skill of observation must be taught to preservice educators throughout their teacher education curriculum, including early experiences during which students observe and measure selected teaching skills (Behets, 1993; Metzler, 1986). In teacher education programs, it is not practical for every student to reach a reliability standard that may take up to forty hours of training in addition to other coursework and field experience hours (Deng Keating, 1999). It is important to determine the amount of training and assistance preservice teachers need to accurately and effectively use systematic observation techniques to optimize learning experiences during their preservice education.

While the measurement of teacher effectiveness has been examined extensively in physical education using a variety of systematic observation instruments, there has been little research looking at the amount and type of training necessary for novice preservice physical educators to accurately code teaching behaviors using computerized systematic observation tools. Furthermore, the research that has been completed has not examined physical education preservice teachers' ability to systematically code teaching behaviors using computer systematic analysis software such as BEST.

Hypothesis

It was hypothesized that participants who receive greater amounts of training would increase their ability to code selected variables using the BEST software system.

Chapter 2

Review of Literature

Introduction

The purpose of this study was to investigate the differences among training protocols used to train preservice physical education teachers as first time users of the BEST software system. In this chapter, literature relevant to the study will be reviewed in the following sections: (a) studies with systematic observation in general education, (b) studies with systematic observation in coaching, (c) studies using systematic observation in physical education and (d) systematic observation studies with computer assistance.

Studies with Systematic Observation in General Education and Supervision

In a study that investigated increasing teacher attention to desired child responses by providing the teacher with factual feedback related to attending behavior, but not providing specific training in reinforcement principles, it was reported that a simple but consistent training procedure could modify teacher behavior, specifically, attending to appropriate child responses (Cooper, Thomson, Baer, 1970).

The purpose of this study was to attempt to increase teacher attention to desirable child responses by providing the teacher with factual feedback related to attending behavior, but not providing specific training in reinforcement principles. Participants were two teachers from different preschools. The participants were in low-income districts of a large midwestern city. Both participants had college degrees and had taught previously in Head Start programs.

The participants had no formal training in reinforcement principles and were observed for eight days to obtain a baseline measure of reinforcement of desirable child responses. Teacher A, who displayed a lower baseline rate of attending to appropriate child responses, was trained first. During training, Teacher A received feedback which included each of the four types of feedback: (a) behavior definition of appropriate child response, (b) local success frequency or number of times attended appropriate child responses, (c) daily rate spent attending to appropriate child responses, and (d) failure frequency or the number of times teacher failed to attend children engaged in appropriate responses. Teacher B was then trained in a similar way. Teacher B was simply observed during the first part of the training condition for Teacher A.

The observer made a written record of teacher behavior every 10 seconds on recording forms. The observer recorded whether appropriate child responses had occurred near the teacher (within 6 ft) during that time, and if so, whether the teacher attended to them. Results indicated that Teacher A's attending rate rose 30% while Teacher B's attending rate rose 14% when they received local success frequency and daily rate spent attending to appropriate child responses. It should also be noted that each teacher's rate began to increase after training was implemented. Both teachers demonstrated an increase in attending to appropriate child responses following the onset of experimental feedback.

Another study examined the effect of two school principals' observation and intervention procedures on the teaching behaviors of three physical education teachers (Ratliffe, 1988). During the first intervention procedure, one principal was asked to conduct a standard observation procedure with the physical education teacher. For these two principals, standard observation procedure meant watching the class and jotting down things to discuss later. This resulted in at least 20 minutes of observation time in two classes and at least one session to share information with the teacher and to make recommendations. For the second intervention, the second principal viewed an instructional videotape and read the companion manual, then observed the teacher for at least 20 minutes in two separate classes and conducted one session to share information and to make recommendations.

The videotape instructed the principal in the use of two systematic observation instruments for observing the teacher. The instructional videotape was designed to demonstrate to the principals what to look for and how to collect objective information about specific teacher behaviors related to management and student activity time. The videotape depicted specific situations and examples of management and student activity time, modeled by the investigator using a class of 3rd and 4th grade students. One instrument focused on class management and was split up into four categories: (a) starting class, (b) getting equipment, (c) giving directions, and (d) changing activities. For ten minutes the principal focused on observing the teacher and the class. The principal used a stopwatch to record the elapsed time for the appropriate category and put either a check in each category to show the teacher demonstrated that behavior or a minus to show the teacher should improve in that specific category.

The other instrument focused on the use of students' time and was split up into six different categories: (a) performing motor activity, (b) receives information, (c) gives information or assists, (d) waits, (e) relocates, (f) off-task behavior. Similar to the class management tool, the principal used a stopwatch to record the elapsed time for the appropriate category and placed either (a) a check in each category to show the teacher demonstrated that behavior or (b) a minus to show the teacher should improve in that specific category. For the third intervention, the physical education teacher viewed the videotape, reviewed the observation instruments, and read the manual. Under the baseline condition, no information was given to the principals or teachers.

Data obtained from the principals' observations were shown to the teacher during the sharing session, with discussion then focusing on reducing management time and increasing student activity time. Results indicated that an increase in student activity time and a decrease in management time did occur after the second and third intervention procedures. The combined effect of all interventions on management time for Teacher A was a reduction of 43% from baseline. Teacher B had more room for improvement and reduced management time 57.2% from baseline. Management time for both teachers remained the same or increased after the standard observation procedure, but decreased after each subsequent intervention.

Studies With Systematic Observation in Coaching

In a study that investigated the coaching behaviors of more and less successful high school boys tennis coaches during practice sessions, it was reported that a good deal of time was spent in management, silence, and other behaviors, which are not usually recognized as productive teaching strategies (Claxton, 1988).

The purpose of this study was to systematically describe and analyze the coaching behaviors of more and less successful high school boys tennis coaches during practice sessions. Participants were five tennis coaches with a 70% or greater career win record and 70% wins in the last three years, and four tennis coaches with less than 50% career win record and less than 50% wins in the last three years.

Data were collected using the Arizona State University Observation Instrument (ASUOI). The instrument used a 14 category coding system: (a) pre instruction, (b) concurrent instruction, (c) post instruction, (d) questioning, (e) manual manipulation, (f) positive modeling, (g) negative modeling, (h) first name, (i) hustle, (j) praise, (k) scold, (l) management, (m) silence, and (n) other.

The nine coaches were observed three times each, once during pre-season, once during mid-season, and once late in the season. Each observation consisted of three 10-minute periods spaced 10 minutes apart for a total of 90 minutes of observation per coach. Trained observers standing on or near the court recorded the data live. Each occurrence of the 14 behaviors on the ASUOI were recorded and behaviors lasting over five seconds were recorded again but marked with a dash to indicate it was a continuation. Five interobserver agreement checks were conducted, producing agreements of at least 80% on all occasions.

Data were analyzed by computing each behavior category into percent of total behaviors. Results indicated that a total of 4,031 discrete behaviors were recorded in 810 minutes of observation. The 4,031 events were depicted by rates per minute and percentages. Concurrent instruction and post instruction combined to account for 20.1% of all behaviors, making instruction the largest single category. Almost 15% of all behaviors were considered in the "other" category and 13.5% were management. Silence accounted for 13% of the behavioral events. Together, "other", "management", and "silence" made up 41.5% of all observed behaviors. Manual manipulation was the least with 0.3%. The praise category represented 9.5% of time spent while the scold category represented 1.8% of time spent.

Interestingly, the less successful coaches instructed more than the more successful coaches. Also, praise was used more by the less successful coaches (12.1%) than by the more successful coaches (7.2%). The only statistically significant differences between more and less successful coaches were questioning (7.4%) to (3.0%) respectively at the .05 level of confidence.

A second study examined the teaching / coaching behaviors of winning high school head football coaches during practice sessions. The purpose of this study was to investigate the teaching / coaching behaviors of winning high school head football coaches during practice sessions (Lacy & Darst, 1985).

The participants in this study were 10 high school head football coaches in AAA classification (minimum 1,600 pupil enrollment) schools in Phoenix, Arizona. Each participant was required to have at least four years experience as a head football coach at the varsity level and .600 or higher career winning percentage.

Data were collected using event recording, which is a cumulative record of the number of discrete events occurring within a specified time. Each time a predefined behavior was observed, that behavior was recorded on the coding sheet. Each practice segment was timed to the nearest minute for the purpose of determining the rate per minute (RPM) of each behavior category occurring during that particular part of the workout. In order to observe and analyze the behaviors of the head coach during specific parts of the workout, practice segments for this study were described as follows:

(a) Warm-up (Included stretching, calisthenics, isometric exercises, footwork/agility drills), (b) Group (Separating the teams into position specific groups), (c) Team (Incorporated game-like situations which all 11 members worked together), and (d) Conditioning (Various forms of running to improve muscular and cardiovascular fitness). There were 11 behaviors examined in this study modified from Tharp and Gallimore (1976). They consisted of: (a) use of first name, (b) praise, (c) scold, (d) instruction, (e) hustle, (f) nonverbal reward, (g) nonverbal punishment, (h) positive modeling, (i) negative modeling, (j) management, and (k) other.

Data were analyzed using a Fortran computer program to perform the quantitative analysis of the observed coaching behaviors. Analysis of variance with repeated measures was used to statistically determine if significant differences existed at the .05 level of confidence between the means of the various coaching behavior categories in the different phases of the season. Results indicated that four of the eleven mean RPMs were significantly different between phases. The four RPMs were the behavior categories of praise, scold, instruction, and positive modeling. In each of the four behaviors (warm-up, group, team, conditioning) a significant difference occurred at the .05 level of confidence between the preseason phase and both the early and late season phases. Most behaviors exhibited throughout the season occurred in either the group segment or the team segment. The group segment accounted for 42.4% of total behaviors and the team segment totaled 45.5%.

The total RPM was higher in the group segment, (5.48) than in any other segment. The team segment RPM was 3.78, followed by the warm-up RPM of 3.05 and the conditioning RPM at 2.93. The instruction category dominated the group and team segments and accounted for 42.5% of all behaviors during the season. The total RPM for all behaviors was higher in the preseason (5.31) than in either the early (3.70) or late season (3.67) phases.

A third study explored the percentage of time in which school pupils coached by teachers were engaged in moderate to vigorous physical activity (MVPA) during extra-curricular sport practices (Curtner-Smith, Sofo, Chouinard & Wallace, 2007).

Participants included 20 high school teachers from Alabama that coached high school girls basketball. Data were collected using the System for Observation Fitness Instruction Time (SOFIT). Practices were videotaped and the verbal behaviors of teachers were recorded using a microphone. Videotapes were coded using SOFIT. Twenty practices were videotaped at three different times (Total=60) throughout the season. Practices were videotaped during early-season, mid-season, and late-season. Practice time averaged 91.45 minutes in the early season, 88.52 minutes in the mid-season, and 71.75 minutes in the late-season. In addition, three target pupils were randomly selected for videotaping during each practice. Target pupils were videotaped during 1-minute intervals in a repetitive rotational order throughout each practice. Interobserver reliability was checked using procedures recommended by Van der Mars (1989). This involved the second and third authors coding a videotaped practice designated as the reliability practice before coding of the study practices began. Both observers compared their results in order to establish inter-rater reliability. Reliability percentages resulting from this check were 86.90% (pupil activity), 91.30% (practice content), and 82.60% (teacher behavior), which exceeded the 80% level recommended by Van der Mars (1989).

Data were input into Statistical Package for the Social Sciences (SPSS) version 10.0 in order to produce descriptive statistics across all 60 practices and for the 20 early, mid, and late season practices. One-way repeated measures analysis of variance (ANOVA) tests were performed with Bonferroni adjustments if necessary, in order to determine whether percentages of time spent in various activities, practice contexts, and teacher behaviors changed during the course of the season. Level of significance was established as $p < .05$.

Results indicated during the course of the season players spent 50.47% of their time engaged in MVPA. Much of this time was spent in very active behavior (31.51%) while the remainder was spent walking (18.96%). Results further indicated players spent 42.30% of their time standing. The amount of time focused on teaching skills and strategies of basketball was 88.01%. Teachers allocated very little time for management (7.55%). The teacher behavior section revealed that teachers main priority was teaching the game of basketball with 75.41% of their time providing pupils with instruction on skills, strategies, and tactics.

A fourth study involving coaching investigated John Wooden and his coaching behavior (Tharp, & Gallimore, 1976). Data were collected using a pencil and paper 10-category systematic observation system. The ten categories coded were: (a) instruction, (b) hustle, (c) modeling-positive, (d) modeling-negative, (e) praises, (f) scolds, (g) nonverbal reward, (h) nonverbal punishment, (i) scold / reinstruction, and (j) other. There was an eleventh category for behaviors that were uncodeable. Two observers collected data live during fifteen practice sessions. Observer agreement was above 90% in all categories.

Results indicated a total of 2,326 acts of teaching were classified into the 10 categories. Instruction constituted 50.3% of Coach Wooden's teaching acts. Total positive social reinforcements, verbal and nonverbal, constituted less than 7% of total acts while scolds added up to 14.6% of total acts. The scold / reinstruction category constituted for 8% of total time while modeling-positive represented 2.8% and modeling-negative represented 1.6% of total time spent. Hustle constituted the most communication next to instruction at 12.7%.

Studies Using Systematic Observation in Physical Education

In a study that examined the effects of a sequential behavior feedback protocol on the practice-teaching experiences of undergraduate teacher trainees, it was reported that an effective teacher education practice was to implement practical experiences guided by sequential behavior feedback which focused on the link between teacher practices and challenging pupil situations in the gymnasium (Sharpe, Lounsbery, Bahls, 1997).

Participants were two male and female undergraduate students enrolled in a junior-level physical education methods class (N=14) who served as participants in the following semester practice-teaching experience. Participants were selected in hopes of (a) making the participant sample as representative as possible of K-12 physical education preservice teachers, and (b) limiting the potential experimental confounds of exposure to teacher education experiences outside of the undergraduate physical education core. The criteria for participant selection included that they had completed certification core coursework, maintained the required 3.0 average or better for all undergraduate work and were scheduled for their elementary practicum and culminating student teaching experience the following year.

Data were collected using a 15-point Likert scale based on information taught in a methods class. A multiple-baseline-across-participants design was used to determine the effects of four feedback protocols which were as follows: (a) exposure to qualitative feedback only on teaching practice (baseline), (b) exposure to sequential behavior feedback on teaching practice (behavioral feedback), and (c) exposure to a feedback on teaching withdrawal phase (maintenance) after exposure to conditions (a) and (b).

Feedback was delivered once per week during the behavioral feedback phase of the study and consisted of 15 minutes of university supervisor and undergraduate participant discussion of sequential behavior data. During baseline conditions each participant received only general qualitative feedback related to teaching performance based on a 15-point Likert scale, with items such as "provided materials at an appropriate level of difficulty for pupils" and "provided a well managed and organized classroom." Qualitative feedback was held to a 15-minute session once per week in which the university supervisor, teacher supervisor, and undergraduate participant discussed their respective perceptions of the strengths and weaknesses of the lesson observed. In addition, during this session they went over the supervisor's Likert scale ratings for that lesson as well as discussed supervisor recommended goals based on the Likert scale information to be implemented for the next practice teaching episode.

During the sequential behavior feedback condition, each undergraduate participant received specific feedback related to the data describing the sequential teacher and pupil behavior patterns for that days teaching performance. Sequential feedback was provided immediately after each practice episode for each participant once per week.

During the maintenance condition, each undergraduate participant was reassigned to a different teacher within the same four settings used in the sequential behavior feedback condition. In this phase the undergraduate participants continued to teach a similar group of pupils for two complete class periods per week and received no post class feedback related to their teaching performance.

Results indicated that qualitative feedback alone in the context of a practice-teaching experience did not promote a high-percentage use of recommended teaching practices in the context of challenging instructional situations. Results further indicated that the effectiveness of qualitative notes in providing feedback on practice teaching performance was minimal. Results further indicated that providing sequential feedback is important to ensure that preservice teachers target gymnasium challenges and deal with those challenges according to recommended educational practices.

The purpose of another study was to quantify behaviors that were associated with high levels of student involvement (Hastie, 1994). Participants included three classes of students in year 10 (15 years old) at a metropolitan high school. Class size averaged 26 students. In addition, three physical education teachers (2 male, 1 female) also participated in this study (Teacher A, B, and C). All teachers had experience with volleyball, coaching sports, and taught the same units at the school for the past 5 years.

Data were collected using a modified ALT-PE tool. Classes were videotaped from a viewing area above the gymnasium. Thirty, 40-minute lessons were observed. Each teacher was videotaped by a color video camera for 10 lessons. A stopwatch was inserted at the base of the screen for each lesson. The microphone on the camera was able to pick up teacher talk. Interobserver reliability checks were conducted on three 15-

minute segments of the lessons for each teacher. Interobserver reliability was calculated at 99% for these segments.

Data were analyzed through descriptive statistics. Mean scores were determined for all ALT-PE categories. In addition, multivariate analysis of variance (MANOVA) was calculated to measure differences between the three classes and teachers for ALT-PE data. These were followed by univariate analyses of variance with the Newman-Keuls technique for post hoc analysis. Pearson product-moment correlations were calculated between teacher variables and student variables to determine any relationships between them.

Results indicated Teacher A had a significantly greater percentage of student engagement in motor appropriate activity than Teachers B and C. The students of Teachers B and C spent significantly more time waiting for turns, being involved in interim activity, and being off-task. Teacher A had the smallest percentage of off-task behavior yet spent the most time in management. Teacher A spent more time directly interacting with students in terms of giving information about the task, where as Teacher C spent more time observing students.

The "observe-concurrent instruction cycle" consisted of teachers reinforcing key points to students through a short intervention during which the teachers would stop the activity to make a correction, followed by a short period of observation. Results indicated that Teacher A would stop games and scrimmages to give feedback about specific plays and would also set expectations for levels of effort and performance. Teacher A also provided concurrent instruction about positioning or skill execution during rallies and would intervene to give feedback on technique or tactics at the end of rallies.

Teachers B & C were less effective teachers and spent considerable time involved in an observe-officiate cycle consisting of watching students with occasional instruction such as, "Go back the other way" or "Change over." There was no information about performance outcomes, nor was there any attention to the key points of skill execution. Teacher A spent only 3% of time in this cycle while Teacher C spent 85% in this cycle.

Results further indicated the more effective teacher in this study had lessons that were characterized by a pattern of interaction with students, involving frequent concurrent instruction, a large number of intervening interactions, and a few periods of observation (all consistent with promoting involvement). This study reported that specific teaching behaviors lead to significantly greater lesson involvement by students in high school physical education classes. It has also confirmed that concurrent instruction was associated with higher amounts of ALT-PE in more effective teaching-learning physical education environments

Another study investigated academic learning time expended by elementary and secondary school students in regular physical education classes (Godbout, Brunelle, & Tousignant, 1983). The purpose of this study was to determine how much academic learning time was experienced by elementary and secondary school students during regular physical education classes. Participants included a total of 61 physical education teachers of both sexes working at both the elementary level (n=30) and at the secondary level (n=31) in the Quebec school system.

Participants were selected through a two-step sampling procedure adapted to the setting of the Quebec school system. Participants at the elementary level were selected by identifying 17 school boards within a radius of 70 kilometers from Quebec City. Fifteen draws were made after an initial weighting system based upon each school system. A total of 11 different school boards were selected. For each selected school board, a list of physical education teachers working with grades four or five was obtained and the appropriate numbers of teachers were randomly selected from the list with an equal number of potential substitutes obtained in the same manner. At the secondary level, a similar selection procedure was used.

Data were collected using the ALT-PE instrument. Coding categories for the ALT-PE instrument were split into three levels of decision for ALT. Level one was type of content and was divided into general content and physical education content. General content was further divided as follows: (a) wait, (b) transition, (c) management, (d) rest, and (e) non-academic instruction. Physical education content was further divided as follows: (a) single skill, (b) sequential skill, (c) competition, (d) fitness, (e) other motor activity, (f) knowledge development, and (g) social development. Level two coded a students' behavior and was divided as follows: (a) engaged, motor response, (b) engaged, compatible motor response, (c) engaged, indirect, (d) engaged, cognitive, (e) not engaged, interim, (f) not engaged, waiting, and (g) not engaged, off-task. Level three indicated the level of students' performance and was divided as follows: (a) succeeds easily, (b) succeeds with some difficulty, and (c) succeeds with great difficulty or fails.

A team of two ALT observers observed each class. At the very beginning of the chosen class period, each observer selected at random three target students, making sure the other observer did not select him or her. Each target student was observed in turn throughout the class period so that each of the selected students was likely to be observed. Approximately two months after their first visit, the observers contacted the same teacher and made arrangements for a second visit with the same procedures being followed. For each observation session, a frequency count was done to determine how many times each of the ALT categories had been coded. Five composite scores were also computed by adding together various categories. Data were analyzed using a t-test for correlated means and the stability of the individual results over time were analyzed through Pearson product-moment correlation coefficient.

Results indicated on the average, that class groups spent from 65% to 81% of the class period in some form of physical education content activity. At this decision level, the wait, rest, and management categories seem to have had more weight at the elementary level. When students were effectively engaged in physical education content activities, they had a very high ratio of success as judged by the observers. This was true on the average for both elementary and secondary schools (with percentages higher than 90%) and was also true for many of the class groups within each level. The resultant ALT-PE figures amounted to averages of 31.3% and 36.5% respectively for the elementary and secondary levels and were found to be significantly different at the .05 level of confidence. It was reported that the main difference found between the elementary and secondary level was in the amount of P.E. content activities versus general content activities. The overall impression was that there was less time lost, at the secondary level, in waiting, managing, and resting, and that this time available was used to increase the competition time in the classroom

Studies Using Systematic Observation with Computer Assistance

Using computerized systems to assist in data collection and analysis procedures in systematic observation has become increasingly popular with the practicality of portable computer hardware as well as the incorporation of more advanced software features (Carlson & MacKenzie 1984; Darst et al., 1989; Kahang & Iwata, 1998; Sidener, Shabani, Carr, 2004). These computerized systems aid in systematic observation and can limit human errors by improving the efficiency of data calculation and graphing while also improving the accuracy of recording. Previous methods of data collection required a shift of attention from the teacher being observed to the recording sheet to enter data. Additionally, data had to be manually recorded and calculated which can be time consuming and can lead to human error. Although there is still room for human error using computer technology, the capabilities using computer systems to compute and analyze data are far more advanced than traditional pencil and paper methods.

The purpose of one study was to compare the data produced by the previously validated and often used System for Observing Fitness Instruction Time (SOFIT) instrument with a computerized instrument, the Computer-SOFIT (C-SOFIT) (Deng Keating, 1999). Participants included fifteen middle school physical education classes selected from a database of videotaped physical education classes. Eight physical education teachers taught the classes. Seven of the teachers taught two classes and one teacher taught a single class.

To collect data, participants coded 15 videotaped physical education classes using both the SOFIT and C-SOFT instruments. This took place in a physical education pedagogy lab. All the observation and recording procedures, other than the instrument, were identical. From the videotapes used for each class, four children were randomly selected as they entered the gymnasium. The focus was rotated among the students every 4 minutes during the coding of the classes. Students wore numbered pinafores so they could easily be identified by the videotape. Data were analyzed by converting all interval data generated by SOFIT into continuous data before data analysis was completed. Intraclass correlations were calculated by analysis of variance to examine the consistency between data collected by the SOFIT and C-SOFIT instruments. Dependent t-tests were calculated between scores generated from both instruments for each of the student activity, lesson context, and teacher behavior categories.

Results indicated there were no differences for any of the student activity or lesson context categories. Results from this study suggest that the C-SOFIT instrument is a viable alternative for data collection focusing on physical activity related instruction in physical education. In addition, it was reported that reliable and valid scores could be obtained from a computerized version of the SOFIT instrument. This study suggested that using computers could enhance the process of data collection in physical education.

Another study used pedometers to quantify physical activity time for first and second grade physical education students (Scruggs, Beveridge, Eisenman, Watson, Shultz, & Ransdell, 2003). Participants were 410 first and second graders in 15 intact classes from six schools in a single school district in the Southwestern United States. Of the 410 students enrolled, 369 received parental consent. Two thirds of the total sample was randomly assigned to the validation sample (n=246). The cross-validation sample consisted of 123 first and second graders.

Data were collected using the C-SOFIT instrument and Yamax Digi-Walker SW 701 pedometer. Before implementation of physical education lessons, participant's stature and body mass were measured. Stature was measured without shoes to the nearest centimeter using a standard 1.83m carpenter's ruler. Body mass was measured without shoes to the nearest kilogram using a commercially purchased electronic scale. Data were also collected through videotaped observations. Video cameras were placed at opposite corners of the gymnasium. In each of the 15 intact classes, participants' activity levels were analyzed once via videotape by trained researchers. Approximately one-third of the participants per intact class were videotaped in one of the three physical education lessons. Colored jerseys were used to identify each participant for later video analysis.

Data were analyzed for each observation session. A frequency count was done to determine how many times each one of the 22 specific categories had been coded. Five composite scores were also computed by adding together various specific categories. The categories were: (a) general content, (b) physical education content, (c) student engagement, (d) student non-engagement, and (e) student success. This was done for each student observed during the class period and the frequencies were summed over all the observed students. The ratio of the final frequency count for each category, or composite score, over the total number of observation intervals (once multiplied by 100) yielded a percentage of class time devoted to a given group of categories.

Results indicated mean lesson time for all 45 lessons was 29.48 ± 1.93 minutes. Mean lesson times for the three-lesson unit were 29.43 ± 1.67, 29.15 ± 2.11, and 29.82 ± 2.08 for lessons 1, 2, and 3 respectively. Mean total steps for each lesson were 1892.33 ± 311.22, 1896.06 ± 309.14, and 1793.96 ± 382.19 for lessons 1, 2, and 3 respectively.

Steps per minute were strongly correlated with percent MVPA (moderate to vigorous physical activity) in both validation and cross-validation samples indicating that physical education classes can be assessed via pedometry in terms of meeting time requirements in physical activity.

The following study attempted to validate the estimates of time spent in various physical activity intensities obtained with the paper and pencil versions of SOFIT during actual physical education classes using the BEST software system which is a computerized system of recording and time keeping (Heath et al., 2006). Participants included one hundred forty-eight third, fourth, and fifth grade boys (n=74) and girls (n=74). Participants were observed during physical education classes at five elementary schools in Cache County, Utah and two schools in El Paso, Texas in fall 2000, spring 2001, and fall 2001. A total of 12 third, 12 fourth, and 13 fifth grade classes were observed (N=37). Consent was obtained from the school districts, principals, and physical education instructors to observe classes as they were conducted. Consent was also obtained from parents to allow their children to participate, if selected.

Third-grade lessons were approximately 30 minutes long and forth and fifth grade lessons were approximately 45 minutes long. The BEST software was loaded onto a laptop computer and programmed specifically for the SOFIT activity lesson and context codes. Observers positioned themselves at a distance where they did not disrupt the instruction but could clearly observe student activity. Data were collected by a random selection of five students from each class (2 boys, 2 girls, and an alternate child) and observing each for 4-minute intervals on a rotational basis. The "paper and pencil" SOFIT version and the Behavioral Evaluation Strategies and Taxonomy (BEST) were used to record children's activity levels simultaneously. Both observers did not change their recording method or timing device and they both observed the same children and made observations at twenty-second intervals throughout the study.

The unit of analysis for all statistical comparisons was the physical education class. Data were analyzed by totaling the number of seconds recorded by BEST and the paper and pencil SOFIT in each of the following activity categories: (a) lying down, (b) sitting down, (c) standing, (d) light activity, (e) moderate activity, and (f) vigorous activity. For the paper and pencil SOFIT the total number of seconds was estimated by multiplying the percentage of observations for each activity category by the total amount of observation time per physical education class.

Results indicated significant comparisons between paper and pencil SOFIT methods and BEST methods of systematic observation. Effect sizes for the differences between the paper and pencil SOFIT methods and BEST were small. Mean scores in seconds were compared between the paper and pencil method and BEST. No significant differences were found in time spent in various intensities of activity between the paper and pencil version of SOFIT and computerized BEST. Results indicated excellent agreement between the paper and pencil method of SOFIT and the computerized BEST version.

Another study explored the effects of computer-assistance during systematic observation on the attitudes of pre-service teachers toward systematic observation and on time required to analyze the data (Pastore, & Peck, 1994). Participants were 36 volunteers that were enrolled in a secondary education pre-student teaching field experience program. The participants were completing a seventh semester practicum course in which they were required to teach a minimum of ten lessons in public schools. The schools utilized in this program were located in central Pennsylvania and consisted of five high schools and one junior high school. Eighteen participants were randomly assigned to the group using computers while the other 18 participants were assigned to the group using "pencil and paper" methods.

Data collection was conducted using a systematic observation instrument that was devised and validated with the guidance of two content experts. The instrument and its instructions were delivered in the form of printed materials and measured the follow-up categories of student and teacher verbal behaviors which were: (a) teacher statements, (b) teacher statements of praise, (c) teacher questions – low inference, (d) teacher questions – high inference, (e) student questions, (f) student statements, (g) wait time one, and (h) wait time two. The computer group used a HyperCard-based program specifically designed to gather data in these categories and to perform appropriate calculations. The "pencil and paper" group used a printed form that had been used to summarize systematic observation data in previous courses.

Participants were required to teach ten lessons during a five-week period and analyze an audiotape of their third and seventh lessons using a systematic observation method. Participants were then randomly assigned to the computer group or the "pen and pencil" group. All participants in both groups used recorders with headphones to avoid distraction during coding. The final report included the total number of observations in each category, the percentages of statements in each category, and the total percentage of teacher talk and student talk.

At the conclusion of the second session, students were given a 10-item survey that assessed attitudes toward systematic observation and computer-assisted systematic observation on a five-point, Likert-type scale. The categories on the 10-item survey were as follows: (a) systematic observation is time-consuming, (b) I feel systematic observation can help me become a more effective teacher, (c) systematic observation requires too many calculations, (d) as a teacher, I would use systematic observation, (e) systematic observation would be easier to use with a computer, (f) systematic observation is useful for analyzing teacher behaviors, (g) I would prefer to use systematic observation on a computer rather than with a pencil, paper, and calculator, (h) systematic observation requires too much equipment, (i) I prefer not to use systematic observation, and (j) systematic observation is too much work.

Data were analyzed using descriptive statistics. A t-test analysis was applied to the results of the attitude survey to assess the differences between groups with the level of confidence set at .01. Results indicated that both groups favored the use of systematic observation and did not believe that it was a tedious or time-consuming process. The computer group did have a more favorable attitude toward the use of the computer with systematic observation. Significant differences were found at the .01 significance level in two of the survey items. Students believed that systematic observation would be easier to do with a computer and students preferred systematic observation on a computer rather than with a pencil, paper, and calculator. There were also significant differences between groups in time necessary for performing quantitative analysis. The computer group reported an average time of 32 minutes and the "pencil and paper" group reported an average time of 55 minutes. Results indicated that computers can reduce the laborintensive processes associated with systematic observation, such as time and effort required for quantitative analysis

In a study that examined preservice physical educators' perceptions of using the Behavioral Evaluation Strategy and Taxonomy (BEST) software program, it was reported that data provided by the software analysis supplied them with undisputable evidence of their teaching performance (James, 2008). The purpose of this study was to examine preservice physical educators' perceptions of using the Behavioral Evaluation Strategy and Taxonomy (BEST) software program. Participants were 25 preservice physical education teacher education students enrolled in a secondary methods class at a comprehensive college in the North East. Data were collected through formal interviews with 25 participants as well as document data in the form of a reflective paper.

Interview data and document data were analyzed qualitatively through constant comparison. Categories were developed and examined for common elements that ran throughout and tied them together. Themes were then extracted from these categories. Data were then selectively coded for examples that illustrated these themes.

Two main findings were drawn from the analysis. First, results indicated through the use of the software, participants were able to personally identify their strengths and weaknesses. Second, results indicated that participants' perceived that their learning was enhanced through use of the software because it provided them with visual representations of their teaching in the form of several different data charts.

Chapter 3

Methods

The purpose of this study was to investigate the differences among training protocols used to train preservice physical education teachers as first time users of the BEST software system. This chapter describes the procedures used in the investigation of the effects of each training protocol. This chapter explains each of the following: (a) participants, (b) instruments or apparatuses, (c) procedures, (d) experimental design, and (e) data analysis.

Participants

For this study, 33 male and female students were randomly selected from a group (N=56) of physical education teacher certification students who were currently enrolled in a *methods of elementary physical education instruction course* at a midsize college located in western New York. Two participants voluntarily withdrew from the study prior to data collection.

This study was submitted for Category II (Expedited Review) and passed through the Institutional Review Board in December of 2008. Recruitment and permission for all participants was obtained through informed consent, which stated the purpose of the study, the participant's role in the study, described the parameters of the study, and clearly stated that their participation was completely voluntary.

Instruments or Apparatuses

The instrument used in this study was the Behavioral Evaluation Strategies and Taxonomy (BEST) software system. BEST is a program designed for educators and researchers for direct observation data collection and analysis related to educational training and development. Intraclass correlations indicated excellent agreement between the paper and pencil methods of data collection and the computerized BEST version (Heath et al., 2006). The BEST software system was operated using Microsoft Windows with the XP operating system.

BEST Software System

The evolution of software programs to systematically observe teaching events and behaviors has simplified the process of data collection and analysis. The BEST system has been particularly useful in simultaneously recording multiple variables and is an effective way to collect, store, and analyze observational data (Sidener et al., 2004). The BEST software application is divided into two programs, one for the collection of data (BEST Collection) and one for data analysis (BEST Analysis).

The BEST Collection software allows up to 36 different responses to be recorded using the A-Z and 0-9 keys on a standard keyboard. Each of these keys can be edited to suit the needs of an observer. The assigned key-tag names are visible to the user via the onscreen keys. In addition, a second function allowed a text feature to input qualitative data during data collection.

The BEST data collection program provides the capability to record eight types of events or behaviors classified as follows: (a) response frequency (type and amount of feedback), (b) duration (the amount of time a specific behavior such as activity or instruction occurred, which then can be converted into a percentage), (c) intervals (analyzed behavior patterns for a short period of time), (d) time samples (observed group behavior as well as identified student effort, activity, and participation), (e) latency (determined the amount of time it takes for classes to respond to commands or signals), (f) interresponse time (IRT), and (g) discrete trials (Kahng & Iwata, 1998).

After data is collected using BEST Collection, the resulting data file can then be analyzed using the BEST Analysis program. There are several analysis options including: (a) qualitative summary, (b) hierarchical presentation of quantitative information (e.g. frequency, duration, latency), (c) sequential analysis (e.g. z-scores, conditional probabilities), and (d) visual illustrations in the form of tables and graphs. Table and graphing options include: (a) scalable time plot which shows bars that determine duration times and slashes which represent each time a frequency key was pressed, (b) bar charts that record the total amount of frequency for each key pressed, and (c) pie chart distributions with percentages. Additionally, a reliability program allows the comparison of interobserver agreement computing overall agreement and kappa.

Procedures

The purpose of this study was to investigate the differences among training protocols used to train preservice physical education teachers as first time users of the BEST software system. Participants were randomly selected and placed in four groups using a counter-balance method based on grade point average from two classes of a *methods of elementary physical education instruction course* (N=56). The counter-balance design separated students so each group had an even number of high and low GPA students. The same Coding Video (CV) was used for each group to code. Participants were asked to use the BEST software system to systematically observe the CV. Initially, participants examined twelve variables while watching the CV. Eight of the twelve variables were the following behaviors: (a) use of names, (b) specific congruent feedback, (c) general feedback, (d) corrective feedback, (e) positive behavior feedback, (f) negative behavior feedback, (g) demonstration, and (h) question. The remaining four variables measured the duration of each of the following: (a) activity, (b) instruction, (c) management, and (d) waiting.

Experimental Design

Three experimental groups and a control group were used in this study. The three experimental groups received differing levels of training, while the control group received no training. The four total groups were as follows: (a) Control (CG), (b) Training Protocol 1 (TP1), (c) Training Protocol 2 (TP2), (d) Training Protocol 3 (TP3).

Content validity was supported by pilot work. A PowerPoint presentation and the BEST training video were viewed by students from a third *methods of elementary physical education instruction course*. Students were asked open-ended questions about content of the presentation and training video. Minor adjustments such as larger text and visuals were made after the pilot work to increase visibility of both training protocols.

Prior to coding, all groups received a coding sheet with definitions of the twelve variables immediately before viewing the Coding Video. Additionally, all groups received information on how to run the BEST software system from the Windows menu. Participants in each of the groups used headphones to avoid distraction during coding. The participants were also restricted from interacting when participating in the study.

The control group (CG) consisted of nine randomly selected students from a *methods of elementary physical education instruction course* (Section A). These students had no prior training nor had they used the BEST software system.

Implementation of Training Protocol 1 (TP1) took place during another *methods of elementary physical education instruction course*, taught by a different professor (Section B). Training Protocol 1 consisted of seven randomly selected students who received information on systematic observation throughout class sessions, similar to participants from Section A; however, participants in TP1 were provided a PowerPoint presentation that consisted of a brief tutorial prior to coding the CV.

Some information for the PowerPoint (Appendix A) was taken from the BEST website tutorial from www.skware.com. The presentation began with background information on the BEST software system including what it was and why it was used. In addition, information was provided in the PowerPoint about steps for the user to operate the software such as starting and stopping the recording screen, pressing the appropriate numeric and alpha numeric keys for the behavior observed, and an overview of the graphs and charts used for this exercise.

Participants in Training Protocol 2 (TP2) included eight different students randomly selected from methods course Section A. Participants viewed a training video in a laboratory immediately before they watched the CV. The training video was approximately ten minutes long and was split up into two parts. Part I of the training video consisted of eight frequency count behaviors and Part II consisted of four variables based on use of students' time. Part I & II were created similarly, providing a verbally stated definition of the behavior, followed by written statement (on screen) of exactly what students will see (i.e. instructor will provide general feedback by saying "Good job"), concluding with specific video clips from a previous methods of *elementary physical education instruction course*.

The training video was created using Microsoft Movie Maker by the primary researcher in this study. Video clips were provided from a previously videotaped *methods of elementary physical education instruction course* to encourage consistency between the training video and the CV.

Participants in Training Protocol 3 (TP3) were provided with the greatest amount of training. Participants in this group consisted of seven different students from Section B and were involved in TP1 and TP2. After viewing both training protocols, each participant then coded the CV in a pedagogy lab.

Data Analysis

Of the twelve variables initially examined in this study, five feedback variables were totaled into one value, termed "total feedback." Total feedback was a combination of: (a) specific congruent feedback, (b) general feedback, (c) corrective feedback, (d) positive behavior feedback, and (e) negative behavior feedback.

To investigate the differences in each group, a control group (CG) (n=9) and three experimental groups (TP1, TP2, TP3) (n=7, n=8, n=7) were used. Data were analyzed by comparing the mean total feedback coded in each experimental group with the mean total feedback coded in the control group. Data were input into the Statistical Package for the Social Sciences (version 17.0). Descriptive statistics were run to calculate mean, median, mode, standard deviation, range, sum, standard error, skewness, and kurtosis. An analysis of variance (ANOVA) was used to determine what differences existed between the four group means. Dunnett post-hoc adjustments were done to determine what differences existed between the experimental groups and the control group. An unpaired t-test (two-tailed) was used to determine what differences existed between participants who viewed the training video and participants who did not.

Chapter 4

Results

The purpose of this study was to investigate the differences amongst training protocols used to train preservice physical education teachers as first time users of the BEST software system. To address this issue, three research questions were investigated. Research question #1 investigated what differences existed between the control group and the experimental groups in coding the video (CV). Research question #2 investigated what differences existed between participants who viewed the training video and participants who did not. Research question #3 investigated if physical education methods classes provided enough instruction about effective teaching behaviors and training in systematic observation for preservice educators to successfully systematically observe and code teaching behaviors.

Research Question #1:

Research question #1 investigated what differences existed between the control group and each of the experimental groups in coding the CV. The CG in this study was compared to three experimental groups: (a) TP1, (b) TP2, (c) TP3 all who received various levels of training. Each participant viewed the CV independently in a pedagogy lab. The BEST Analysis software system automatically totaled the frequency counts for each participant. Data were input into to SPSS version 17.0.

Initially, descriptive statistics were run using SPSS. Figure 1 demonstrates that the z-scores of skewness and kurtosis are considered to be within acceptable limits of normality ($\pm$ 2.0). Mean frequency count and standard deviation for each group were also determined and can be found in Figure 2. Due to unequal group sizes Levene's test for equality of variance was used to determine homogeneity of variance across groups. Results indicated that equal variance was assumed ($p = 0.376$) and a one-way analysis of variance (ANOVA) found differences between the four groups ($F(3,27) = 2.783$, $p = 0.060$). Dunnett post-hoc tests were run to compare each of the experimental groups to the control group. Statistical tests with Dunnett determined significance levels between the CG and TP1, CG and TP2, and CG and TP3 to be ($p = .284$, .041, and .075), respectively. In order to determine meaningfulness of the treatment, omega squared (ω^2) was used. This statistical test indicated that 25.7% of variance was accounted for by the treatment and this effect is considered to be large (Cohen, 1964 indicated $\omega > .20$ is considered to be large). Effect size was determined for the various post-hoc tests comparing each experimental group with the control group. Effect sizes were determined to be .829, 1.294, and 1.198 for TP1, TP2, and TP3 respectively and all effect sizes were considered to be large.

In summary, it was hypothesized that the experimental groups would code more feedback when compared to the CG with different levels of training. The statistical analysis supported this hypothesis because there were greater increases in total feedback coded by TP2 and TP3 when compared to the CG. This suggests that more training increases the amount of feedback coded by first time users when coding videotape using the BEST software system.

Figure 1 – Z-Scores of Skewness and Kurtosis (ANOVA)

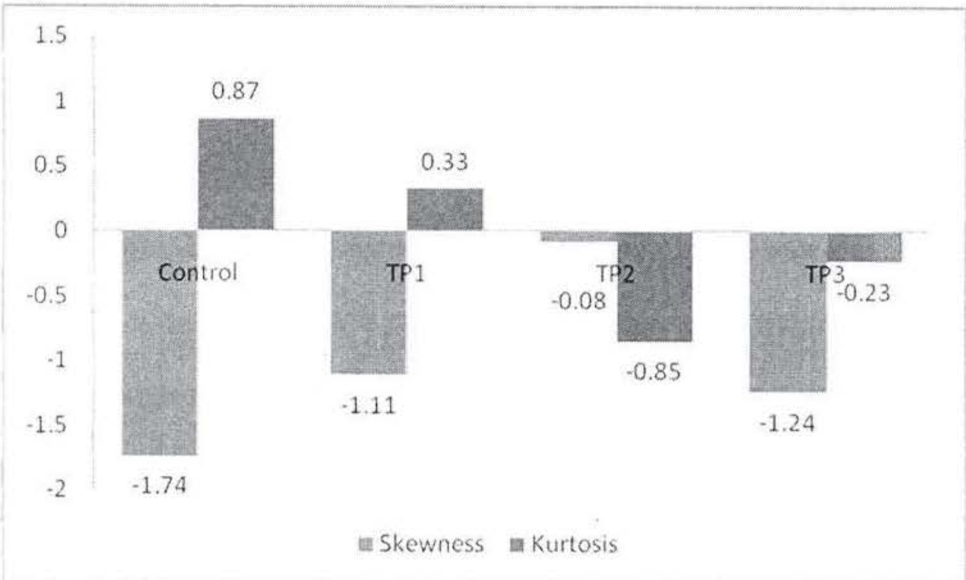
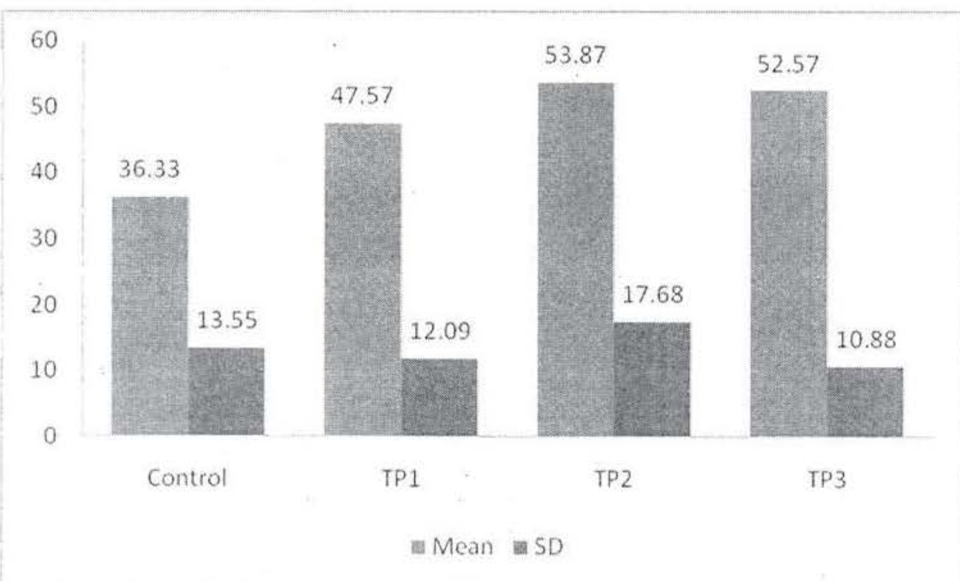


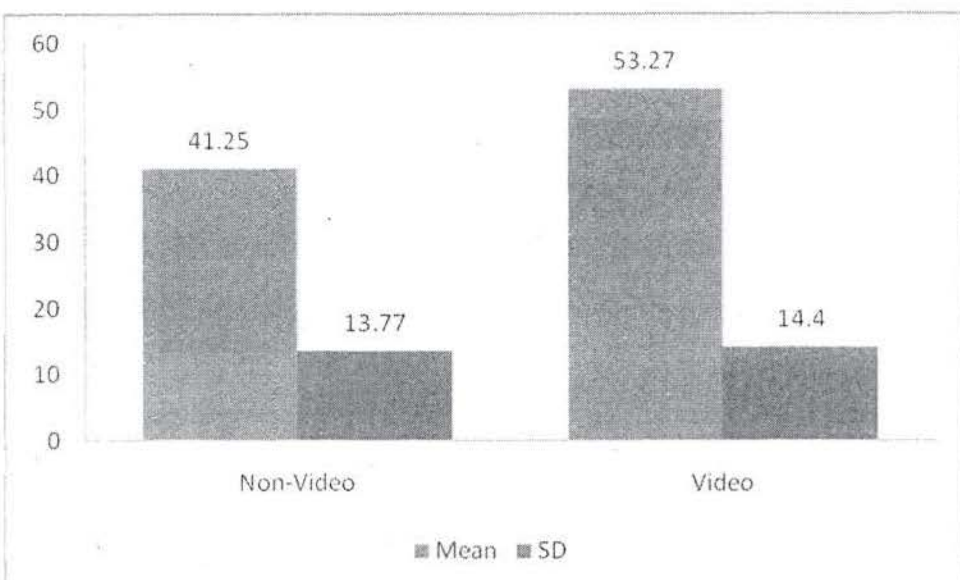
Figure 2 – Mean and Standard Deviation (ANOVA)



Research Question #2:

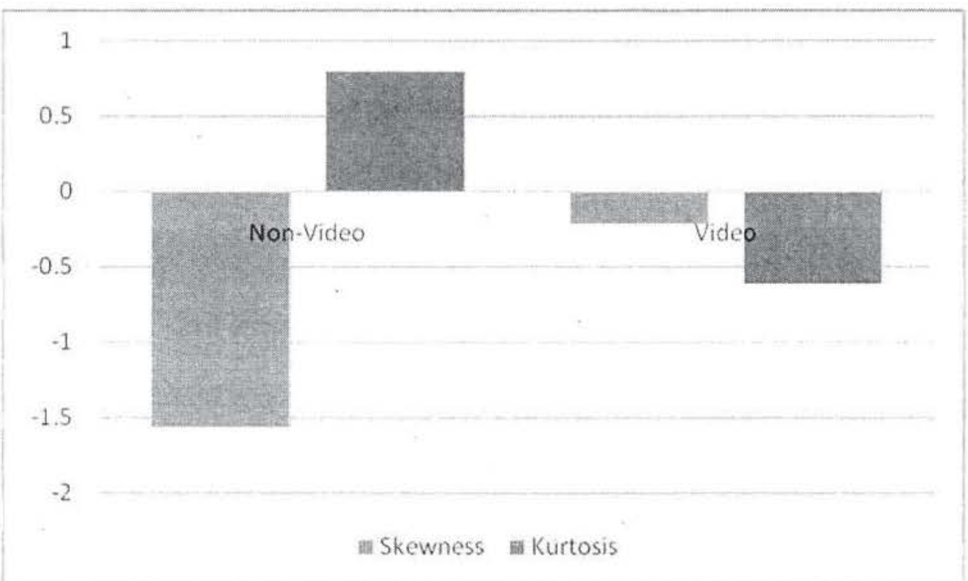
Research question #2 investigated what differences existed between participants who viewed the training video and participants who did not. Although differences (using statistical tests with Dunnett) were found between the control group with each experimental group with significance at $p = .284$, $p = .041$, and $p = .075$ respectively, only TP2 and TP3 involved a training video, while the CG and TP1 did not. To determine what differences existed between the participants who viewed the training video (TP2 & TP3) and participants who did not (CG & TP1), frequency counts of the feedback categories were totaled and compared. The CG and TP1 feedback values were combined into a Non-Video group (NVG) (n=16) whereas TP2 and TP3 feedback values were combined into a Video group (VG) (n=15). Mean values and standard deviation can be found in Figure 3.

Figure 3 – Mean and Standard Deviation (T-Test)



Initially, descriptive statistics were run using SPSS. Figure 4 demonstrates that the z-scores of skewness and kurtosis were considered to be within acceptable limits of normality (± 2.0). Mean frequency count and standard deviation for each group were also determined and can be found in Figure 4. Due to unequal group sizes Levene's test for equality of variance was used to determine homogeneity of variance across groups. Results indicated that equal variance was assumed (p =.612). An unpaired t-test (two-tailed) determined that the differences between participants who viewed the video were greater than the participants who did not ($p = 0.025$). To determine meaningfulness of the treatment, omega squared (ω^2) was used. This statistical test indicated 12.9% of variance accounted for by the treatment. The effect size was determined to be -0.8725, which was considered to be large.

Figure 4 – Z-Scores of Skewness and Kurtosis (T-Test)

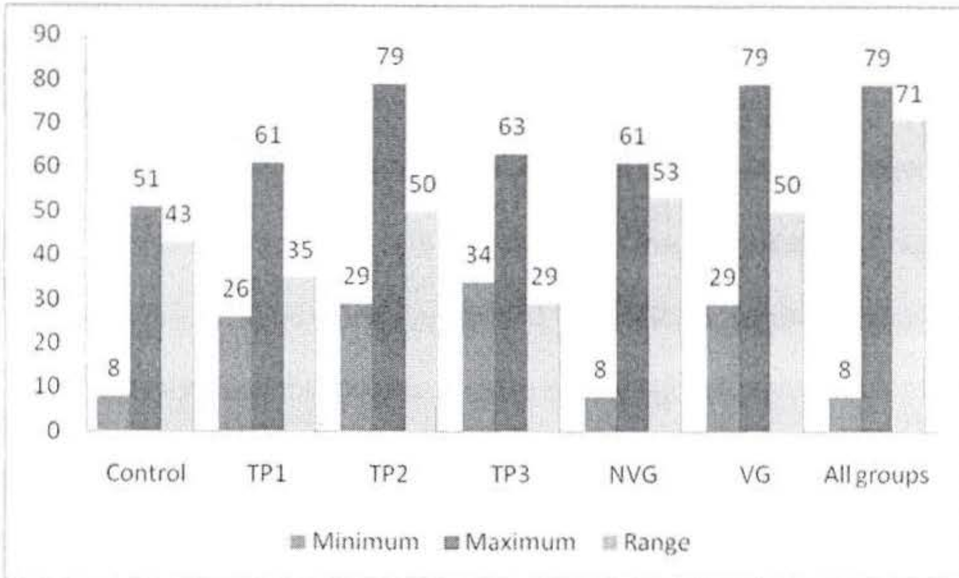


In summary, it was hypothesized that participants that viewed the training video would code more feedback when compared to participants who did not. Statistical findings supported this hypothesis because there was greater total feedback coded by the VG (p =0.025) when compared to the NVG. This would support that viewing specific examples of behaviors being coded increased the amount of feedback coded.

Research Question #3:

Research question #3 investigated if physical education methods classes were providing enough instruction about effective teaching behaviors and training in systematic observation for preservice educators to successfully systematically observe and code teaching behaviors. Initially, descriptive statistics were run using SPSS. Figure 2 indicated the mean feedback coded and standard deviation for each group. The group with the most training (TP3) had the least amount of variability with a standard deviation of 10.88 and a mean comparable to TP2. Figure 3 indicated a slightly larger standard deviation for the video group, although the mean feedback coded for each group was greater. Figure 5 indicates the minimum and maximum amount of feedback coded as well as the range of feedback frequency coded when viewing the CV. A participant from the NVG coded the least amount of total feedback occurrences (8) while a participant from VG coded the most amount of total feedback occurrences (79) while viewing the same video (CV). These three figures indicated a trend between an increased amount of feedback coded and a decrease of the variability of scores when participants were provided with increased levels of training. More specifically, when participants were provided with the training video, they were more consistent in coding greater amounts of feedback behaviors while watching the CV.

Figure 5 – Minimum, maximum, and range of frequency counts



Results Summary

In this study, it was hypothesized that participants who received greater amounts of training would increase the participant's ability to code selected variables using the BEST software system. Results indicated a trend toward an increased amount of total feedback coded when participants were provided with greater amounts training. In addition, the training video used in this study was shown to effectively increase the amount of feedback coded by participants who viewed the CV. It is recommended physical education methods classes provide further instruction about effective teaching behaviors and training in systematic observation.

Chapter 5

Discussion

The purpose of this study was to investigate the differences among training protocols used to train preservice physical education teachers as first time users of the BEST software system. Since there is little information available, there is a need for information regarding the type of training needed for first time preservice educators to be successful using systematic observation, as there is little information available. In a comprehensive literature review, several studies used some type of systematic observation strategy, although none described a specific training protocol to be implemented for first time users. Information that was available reviewed various training attempts with durations of between ten and forty hours of time spent training participants to reach a reliability standard (Behets, 1993; Deng Keating, 1999; Ratliffe, 1988).

In this case, reliability referred to the degree in which two or more observers gave consistent results when viewing the same video. In teacher education programs, it is not practical for every student to reach a reliability standard in addition to other coursework and field experience hours. In addition, research has not provided any information on the ability of first time users to code video using systematic observation techniques. Therefore, there is a need to determine a practical type of training necessary for first time users coding teaching behaviors using systematic observation software (i.e. using BEST to compare various training protocols). It was hypothesized that participants who received greater amounts of training would increase their ability to code selected variables using the BEST software system.

Results based on three research questions will guide this discussion. Research question #1 investigated if there were differences between the control group and the experimental groups in coding the video (CV). Results indicated a trend of students coding more feedback with increased amounts of training. Research question #2 investigated what differences existed between participants who viewed the training video and participants who did not. An unpaired t-test (two tailed) found greater differences $p = 0.025$ for participants who viewed a training video when compared to participants who did not. Research question #3 investigated if physical education methods classes were providing enough instruction about effective teaching behaviors and training in systematic observation for preservice educators to successfully systematically observe and code teaching behaviors. Results indicated a trend between an increased amount of feedback coded and a decrease of the variability of scores when participants were provided with increased levels of training, so it is recommended more training is needed in *methods of elementary physical education instruction courses.*

Research Question #1:

Research question #1 investigated what differences existed between the control group and each of the experimental groups in coding the CV. The CG was compared to three experimental groups: (a) Training Protocol 1 (TP1), (b) Training Protocol 2 (TP2), (c) Training Protocol 3 (TP3) all who received various levels of training. Training Protocol 1 consisted of a PowerPoint presentation. Training Protocol 2 consisted of a ten-minute training video. Training Protocol 3 consisted of PowerPoint presentation from TP1 combined with the ten-minute training video from TP2. The three experimental groups were provided with separate training protocols while the control group was not provided with any training. All four groups were provided with a coding sheet prior to viewing the coding video that provided definitions of the types of feedback being coded (Appendix D).

The first experimental group with the least amount of training (TP 1) consisted of a PowerPoint presentation (Appendix A) provided to the participants in a classroom followed by a brief question and answer session. The Power Point presentation was similar to training protocols used in two studies using the System for Observing Fitness Instruction Time (SOFIT) that investigated the percentage of time students were in moderate to vigorous physical activity (MVPA) (Curtner-Smith et al., 2007; Scruggs et al., 2003).

In the study by Curtner-Smith et al., (2007) observer training involved definitions of SOFIT categories, examples of each pupil behavior, as well as coding full-length videotaped practices. The training protocol was carried out during a one-month period for approximately ten hours, followed by interobserver reliability checks.

In the study by Scruggs et al., (2003) the total amount of hours of training was not mentioned, although participants in this training protocol read SOFIT articles, studied physical activity code definitions, and practiced coding a "Gold-Standard" videotape followed by interobserver reliability checks. In these two studies, coding definitions were provided to each observer that was consistent with the current study that provided a coding sheet with definitions to each participant.

Participants in TP 2 were provided with a slightly enhanced level of training. Each participant in this experimental group viewed a training video that consisted of clips from a previous *methods of elementary physical education instruction course*. The training video depicted each type of feedback by verbally and visually stating the type of feedback and providing a definition followed by giving a visual presentation of a preservice physical education teacher stating the feedback to students in a physical education atmosphere. Participants viewed the training video independently on a computer with headphones to prevent any outside distractions.

This training video was similar to video used in a study by Ratliffe (1988). This study investigated the effects of various intervention procedures on the observation skills of two school principals that observed two physical education teachers. In this study, an instructional videotape was used to demonstrate to the principals what to look for and how to collect objective information about specific teacher behaviors related to management and student activity time. The videotape depicted specific situations and examples of management and student activity time. The principals in this study spent approximately 60 minutes viewing the instructional videotape and practiced using the coding instruments; although the total amount of time spent training was estimated at six hours. Other training involved meetings, observation, and discussion. Results reported by Ratliffe (1988) indicated that training principals to systematically observe specific teaching behaviors in a physical education classroom and then conferencing with the physical education teachers about those observations led to an increase in student activity time and a decrease in management time in the physical education classroom.

Training Protocol 3 was a combination of the first two training protocols and provided participants with the greatest amount of training. Participants were first involved in the PowerPoint presentation (TP1) and also viewed the training video (TP2) before viewing the coding video.

Statistical tests with Dunnett determined significance levels between the CG and TP1, CG and TP2, and CG and TP3 to be $p = .284$, .041, and .075, respectively. The statistical analysis supported the hypothesis that the experimental groups would code more feedback when compared to the CG with different levels of training because there were greater increases in total feedback coded. This result suggests that more training increased the amount of feedback coded by participants using the BEST software system. As revealed by the statistical tests with Dunnett, the largest differences between the experimental groups with the control group were with TP2 and TP3, with the significance levels for training protocols using the training video to be $p = .041$ for TP2 and $p = .075$ for TP3. The training protocol with the most amount of training (TP3) did not result in the largest differences with the control group; whereas Training Protocol 2 did, and involved only the training video. Participants in Training Protocol 2 and Training Protocol 3 viewed the training video immediately before viewing the coding video; however, participants in Training Protocol 3 viewed the PowerPoint approximately two weeks before viewing the coding video. This may explain why more training in TP3 may have not yielded a lower significance level that TP2.

Research Question #2:

Research question #2 investigated what differences existed between participants who viewed the training video and participants who did not. Although differences (using statistical tests with Dunnett) were found between the experimental groups with the control group, the significance levels of the training protocols that used the training video (TP2, $p = .041$ and TP3, $p = .075$) were much lower than that of TP1 ($p = .284$). These results suggest that the training video was most effective in training preservice teachers to use systematic observation. To determine what differences existed between the participants who viewed the training video (TP2 & TP3) and participants who did not (CG & TP1), frequency counts of the feedback categories were totaled and compared. The CG and TP1 feedback values were combined into a non-video group (NVG) (n=16) whereas TP2 and TP3 feedback values were combined into a video group (VG) (n=15).

Results supported the hypothesis that participants who viewed the training video would code more feedback when compared to participants who did not. Statistical findings supported this hypothesis because participants that viewed the training video coded more total feedback ($p = 0.025$) when compared to participants who did not. Results indicated that viewing specific examples of behaviors to be coded increased the amount of feedback coded by first time users of BEST.

The video was a superior method of training for two reasons. First, the training video was viewed immediately before participants coded the CV, thus providing participants with instruction, and then immediate feedback regarding their performance. Second, the instruction provided in the training video was directly related to the variables being coded, compared to the PowerPoint, which was directed as a tutorial of BEST. The training video provided participants with information on definitions as well as specific examples of each behavior to be coded, whereas the PowerPoint provided participants with instruction on how to use the BEST program.

Research Question #3:

Research question #3 investigated if physical education methods classes were providing enough instruction about effective teaching behaviors and training in systematic observation for preservice educators to successfully systematically observe and code teaching behaviors. Participants coded a wide range of feedback responses (8-79), which implied that participants did not get enough training in methods classes. This finding suggests that there is a need for more training using video with the BEST system during methods classes to create less variability between scores. Furthermore, results revealed a trend between an increased amount of feedback coded and a decrease of the variability of scores when participants were provided with increased levels of training. More specifically, when participants were provided with the training video, they were more consistent in coding greater amounts of feedback behaviors while watching the CV.

Although this trend is interesting, there may be another reason why the result did not reach a certain level of significance (i.e., $p < 0.05$). A statistical program (G*Power 3) was used to determine whether there was adequate power to find significant differences ($p < 0.05$) and if not, what sample size was necessary to find statistical significance at the 0.05 level (Faul, Erdfelder, Lang, & Buchner, 2008). G*Power 3 was used to determine and control for type 1 and type 2 errors based on alpha, 1- alpha, beta, 1- beta, and sample size. Based on post-hoc analysis of the data using G*Power 3, a power (1- beta) of 0.63 was determined. Since a power of 0.80 is generally recommended (Cohen, 1988) a value of 0.63 suggests there was insufficient power to find statistical significance at the 0.05 level with an ANOVA using 4 groups and a total of 31 subjects. G*Power 3 determined that 44 subjects were needed to find statistical significance at the 0.05 level (with an effect size of 0.5) with an ANOVA. This would suggest a type 2 error may have been committed (i.e., acceptance of the null hypothesis when it is false). If the same study were to be replicated, it is recommended that at least 44 subjects be used, to ensure there is sufficient power.

Conclusions

Results suggest that training in systematic observation techniques using BEST need to involve a training video similar to the one used in this study. It was found that the training video increased the amount of feedback coded by first time users of BEST. It is, however, unclear whether the feedback coded was correct or incorrect. In order to ensure participants in teacher education programs using systematic observation are accurate in coding video, perhaps a more prescriptive program should be followed.

For example, a teacher education program could include systematic observation experiences that begin early in preservice teachers' careers. During introductory classes, students could be trained in the use of computerized systematic observation and exposed via video and definitions to effective teaching behaviors. Videotape training, similar to the training used in this study, would help provide training that allows students to visually observe effective teaching behaviors. Students would then have a chance to view effective teaching behaviors and then compare their own teaching via videotape while using the BEST software system.

There are several implications for teacher education programs that can be gleamed from the results of this study. First, participants in this study were taken from a methods class in which they had to code twelve total variables while coding the video. In addition to five feedback categories, participants were attempting to code the frequencies of three other behaviors (use of name, number of demonstrations, and amount of questions asked by the teacher) as well as four variables consisting of how the teacher was using class time (instruction, management, activity, and waiting time).

As first time users of BEST, participants were still trying to become comfortable with using the program. Participants had to identify the feedback while watching the coding video, then recall the defined key on the keyboard, press the indicated key, and focus their attention back to the coding video for twelve total variables. Participants also had to remember which keys were defined as duration recording keys and which were defined as frequency count keys. All factors may have contributed to the variability within groups. As a result of these factors it is recommended first time users of BEST code the five feedback variables that were used in this study.

Second, teacher education programs could create and use a gold standard videotape to facilitate student learning in regard to using the BEST system to code teaching behaviors. The "Gold Standard" videotape could contain a predetermined number of feedback statements. This videotape would act as the coding video for each participant while coding the five feedback categories. Furthermore, different gold standard videotapes could be created focusing on other variables to be used as benchmarks for coding different teaching behaviors throughout the careers of teacher education candidates.

In addition, the teacher in the gold standard lesson should be an expert teacher. Not only would having an expert teacher teach the gold standard lesson result in correct demonstration of the targeted teaching behaviors, but also would substantially decrease the amount of time spent establishing inter and intrarater reliability for the coding video.

As preservice teachers progress through their teacher education program, other learning experiences should be offered that involve coding video (that has been previously coded by experts) to provide additional practice coding that is specific to teaching behaviors. These coding experiences would serve as checkpoints to ensure that preservice teachers are developing skills to accurately code video throughout their educational experience.

Third, it is extremely important for teacher education programs to have the necessary equipment to record each and every event occurring in the gymnasium. A wide lens camera should be positioned at a stationary spot in the gymnasium to record behaviors of instructors and students at all times. If this is not possible, trained camera people (one for each instructor) should be instructed on the necessary components to be recorded throughout the lesson. Each cameraperson must remain focused on the teacher(s) as well as each student in the class the entire class period. In addition, instructors should have a microphone attached to their person. This will ensure that all audible information provided during the lesson can be coded when using systematic observation techniques.

Fourth, teacher education programs should also consider the context in which participants are coding videotape. It is recommended participants use headphones to minimize any outside noise and distractions. In regard to viewing the observed lesson, a television may be used to view the videotape; however, the videotape can be uploaded to a computer so it can be viewed on the same screen as the BEST Collection. This would eliminate the participant changing focus from the television to the computer screen repeatedly during the coding process.

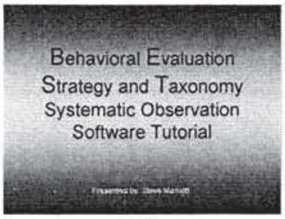
Future Research

Future research should investigate the accuracy of the coding ability of participants when using computerized systematic observation software. Although results of this study indicated that participants coded more feedback when provided with more training, it did not examine how accurate the participants were able to code with the different training protocols. Future research should consider using a gold standard video to assess the accuracy of each participant. Frequency count data from each participant could then be compared to the gold standard video to determine the more suitable method of training.

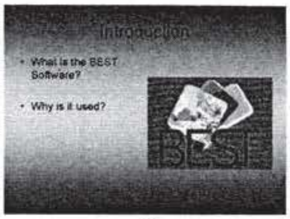
In addition to examining the accuracy of the coding of participants, future research should investigate how reliable participants are when coding systematic observation data. For example, participants could be asked to code a video and then six months later code the same video to determine the reliability of their coding ability. Furthermore, the reliability of preservice teachers' ability to code systematic observation data could be examined after they were exposed to different training protocols.

Finally, research needs to be done to examine the most suitable number of variables to be accurately coded by first time users. Studies could compare the coding ability of participants that viewed one variable, five variables, and twelve variables and determine how accurate they were in coding different numbers of variables. In addition to preservice physical education teachers, the ability of inservice teachers, supervisors, coaches, and administrators to accurately and reliably code specific variables should be investigated.

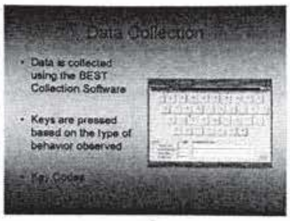
Appendix A



1

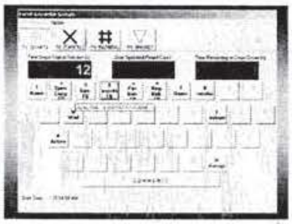


2

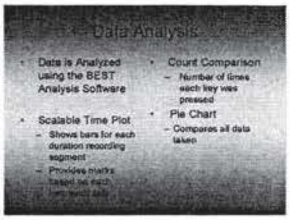


3

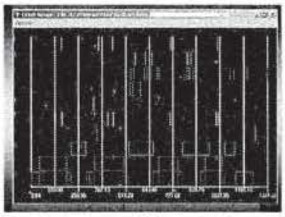
4

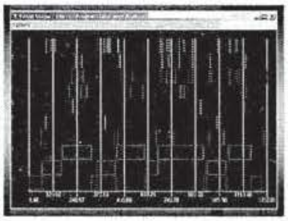


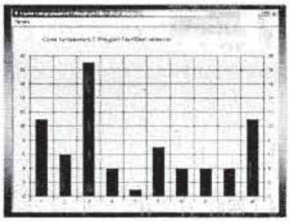
5

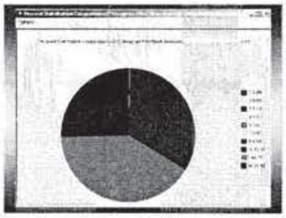


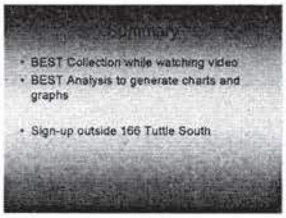
6


7

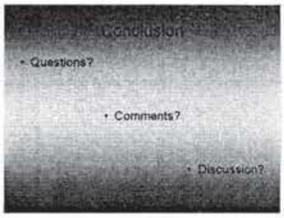

8


9


10



11



12

Appendix B

ANOVA

Descriptive Statistics

	N	Range	Minimum	Maximum	Sum	Mean		Std. Deviation	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Control	9	43.00	8.00	51.00	327.00	36.3333	4.51848	13.55544	183.750	-1.245	.717	1.219	1.400
TP1	7	35.00	26.00	61.00	333.00	47.5714	4.57143	12.09486	146.286	-.878	.794	.530	1.587
TP2	8	50.00	29.00	79.00	431.00	53.8750	6.24911	17.67514	312.411	-.060	.752	-1.265	1.481
TP3	7	29.00	34.00	63.00	368.00	52.5714	4.11071	10.87592	118.286	-.985	.794	-.367	1.587
Valid N (listwise)	7												

ANOVA

TotalFB

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1621.567	3	540.522	2.783	.060
Within Groups	5244.304	27	194.233		
Total	6865.871	30			

Test of Homogeneity of Variances

TotalFB

Levene Statistic	df1	df2	Sig.
1.076	3	27	.376

Multiple Comparisons

TotalFB

Dunnett t (2-sided)[a]

(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
2.00	1.00	11.23810	7.02347	.284	-6.3126	28.7888
3.00	1.00	17.54167*	6.77205	.041	.6192	34.4641
4.00	1.00	16.23810	7.02347	.075	-1.3126	33.7888

a. Dunnett t-tests treat one group as a control, and compare all other groups against it.

*. The mean difference is significant at the 0.05 level.

Appendix C

Unpaired t-test

Descriptive Statistics

	N	Range	Minimum	Maximum	Sum	Mean		Std. Deviation	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
NonVideo	16	53.00	8.00	61.00	660.00	41.2500	3.44299	13.77195	189.667	-.883	.564	.870	1.091
Video	15	50.00	29.00	79.00	799.00	53.2667	3.71800	14.39974	207.352	-.128	.580	-.693	1.121
Valid N (listwise)	15												

Independent Samples Test

	Levene's Test for Equality of Variances		t-test for Equality of Means						
								95% Confidence Interval of the Difference	
	F	Sig.	t	df	Sig. (2-tailed)	Mean Diff	Std. Error Diff	Lower	Upper
Data Equal variances assumed	.264	.612	-2.375	29	.024	-12.01667	5.05979	-22.36509	-1.66824
Equal variances not assumed			-2.371	28.646	.025	-12.01667	5.06731	-22.38606	-1.64727

Appendix D

BEST Analysis Key Codes Defined

Specific Congruent feedback

Feedback that offers usable information specifically related to the task.

General feedback

Informs a learner or group of learners a simplified statement about their skill performance or behavior which follows soon enough after the behavior that the student clearly associates it with the behavior commented on.

Corrective Feedback

This type of feedback informs the learner that their response was incorrect with the knowledge of the correct or desired response.

Positive Behavior Feedback

Instructor makes a positive verbal statement or gesture following an individual's or group of students' skill or organizational behaviors, which are clearly designed to increase or maintain such responses in the future.

Negative Behavior Feedback

Instructor makes a negative verbal statement or gesture following an individual's or group of students' skill or organizational behaviors, which are clearly designed to decrease or eliminate such responses in the future.

Reference:

American Association of School Administrators. (1992). *The nongraded primary: Making schools fit children.* Arlington, VA: Author.

Bailey, D.B., & Wolery, M. (1989). *Assessing infants and preschoolers with handicaps.* New York: Macmillan Publishing Company.

Behets, D. (1993). Systematic observation training for pre-service physical education teachers. *Physical Educator,* 50(2), 87.

Carlson, R.B., & McKenzie, T.L. (1984). Computer technology for recording, storing, and analyzing temporal data in physical activity settings. *Journal of Teaching in Physical Education, 4,* 24-29.

Claxton, D. B. (1988).A systematic observation of more and less successful high school tennis coaches. *Journal of Teaching in Physical Education, 7,* 302-310.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cooper, M.L., Thomson, C.L., & Baer, D.M. (1970). The experimental modification of teacher attending behavior. *Journal of Applied Behavior Analysis, 3,* 153-157.

Curtner-Smith, M., Sofo, S., Chouinard, J., & Wallace, S. (2007). Health-promoting physical activity and extra-curricular sport. *European Physical Education Review, 13(2),* 131-144.

Darst, P., & Pangrazi, R. (2005). *Dynamic physical education for secondary school students.* New Jersey: Benjamin Cummings.

Darst, P.W., Zakrajsek, D.B., Mancini, V.H. (1989). Analyzing physical education and sport instruction. Champaign, IL: Human Kinetics.

Deng Keating, X. (1999). Measuring teaching behaviors, lesson context, and physical activity in school physical education. *Measurement in Physical Education & Exercise Science,* 3(4), 207.

Donat, P. (1991). Measuring behavior: The tools and the strategies. *Neuroscience & Biobehavioral Reviews, 15,* 447-454.

Eiler, J. M., Nelson, W. W., Jensen, C.C. & Johnson, S.P. (1989). Automated data collection using bar code. *Behavior Research Methods, Instruments, & Computers, 21,* 53-58.

Faul, F., Erdfelder, E., Lang , A.G., & Buchner, A. (2007). G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39 (2)*, 175-191.

Godbout, P., Brunelle, J., & Tousignant, M. (1983). Academic learning time in elementary and secondary physical education classes. *Research Quarterly for Exercise and Sport, 54(1)*, 11-19.

Hawkins, R.P. (1982). Developing a behavior code. *Using observers to study behavior: New directions for methodology of social and behavioral science.* San Francisco, CA: Jossey-Bass.

Hastie, P. A. (1994). Selected teacher behaviors and student ALT-PE in secondary school physical education. *Journal of Teaching in Physical Education*, *13*, 242-259.

Heath, E.M., Coleman, K.J., Lensegrav, T.L., & Fallon, J.A. (2006). Using momentary time sampling to estimate minutes of physical activity in physical education: Validation of scores for the system for observing fitness instruction time. *Research Quarterly for Exercise and Sport, 77(1)*, 142-146.

James, A.R. (2008). Technology and systematic observation: Preservice physical educators' perceptions. *Research Quarterly for Exercise and Sport, 79 (1)*, 53.

Johnston, J.M., & Pennypacker, H.S. (1980). *Strategies and tactics of human behavioral research.* Hillsdale, NJ: Lawrence Erlbaum.

Kahng, S, & Iwata, B (1998). Computerized systems for collecting real-time observational data. *Journal of Applied Behavior Analysis*, *31*, 253-261.

Lacy, A.C. & Darst, P.W. (1985). Systematic observation of behaviors of winning high school head football coaches. *Journal of Teaching in Physical Education*, *4*, 256-270.

Maeda, J. (2001). Teacher coaching in physical education: A review. *Physical Education*, 58(3), 140.

McKenzie, T.L., Sallis, J.F., & Nadar, P.R. (1991). Sofit: System for observing fitness instruction time. *Journal of Teaching in Physical Education*, *11*, 195-205.

Pastore, R., & Peck, K.L. (1994). The effects of computer-assisted systematic observation on the attitudes and efficiencies of pre-service teachers. *Journal of Information Technology for Teacher Education, 3(1)*, 91-100.

Ratliffe, T. (1988). Principal training for effective staff development. *Journal of Teaching in Physical Education*, *7*, 228-234.

Scruggs, P.W., Beveridge, S.K., Eisenman, P.A., Watson, D.L., Shultz, B.B., & Ransdell, L.B. (2003). Quantifying physical activity via pedometry in elementary physical education, 1065-1072.

Sharpe, T., Lounsbery, M., & Bahls, V. (1997). Description and effect of sequential behavior practice in teacher education. *Research Quarterly for Exercise and Sport*, *68*, 222-232.

Sidener, T., Shabani, D., Carr, J. (2004).A review of the behavioral evaluation strategy and taxonomy (BEST) software application. *Behavioral Interventions*, *19*, 275-285.

Tharp, R.G., & Gallimore, R. (1976).What a Coach can Teach a Teacher. *Psychology Today*, 75-78.

Thomas, J.R., Nelson, J.K., & Silverman, S.J. (2005). *Research methods in physical activity*. Champaign: Human Kinetics.