AN ABSTRACT OF THE DISSERTATION OF

Jeff McNamee for the degree of Doctor of Philosophy in Exercise and Sport Science

presented on November 21, 2003.

Title: Accuracy of Momentary Time Sampling: A Comparison of Varying Interval

Lengths Using SOFIT.

Redacted for Privacy

Abstract approved: _____   _____   _____
                              Hans van der Mars

The U.S. Department of Health and Human Services has made the promotion

of regular physical activity a national health objective, and experts believe that

physical education can play a significant role in the promotion of physical activity.

Feasible measurement tools to assess physical activity behavior, by physical

educators, are lacking. One validated instrument is the System for Observing Fitness

Instruction Time (SOFIT; McKenzie, Sallis & Nader, 1991). SOFIT's physical

activity data are collected using momentary time sampling (MTS) with a 20-second

interval length and provide estimates of Moderate to Vigorous Physical Activity

(MVPA). Whether variations in interval lengths would adversely affect the accuracy

of the MVPA data has not been investigated. From a clinical perspective, if physical

education teachers are to utilize MTS procedures for on-going assessment they will

require longer time intervals to collect accurate MVPA data. Therefore, this project

sought to determine the accuracy of MVPA levels collected through varying

observation tactics (i.e., 20s, 60s, 90s, 120s, 180s, and random) relative to those collected through duration recording (DR). Video records of 30 randomly selected elementary school physical education classes were utilized for this study. Utilizing modified physical activity codes from SOFIT, the researchers collected MTS data regarding students' MVPA at varying interval lengths (i.e., 20s, 60, 90s, 120s, 180s, and random). Three statistical techniques, Pearson-product moment (PPM) correlation coefficients, Repeated Measures Analysis of Variance (RM ANOVA), and Average Error (AE), were utilized to demonstrate concurrent validity of the varying interval lengths. Results demonstrated moderate-low to high correlations between the 20s, 60s, 90s, and random interval lengths and the DR tactic during the total class. The RM ANOVA indicated similarity between all the varying interval lengths and the DR tactic for total class observation. The MTS procedure that created the least amount of AE across classes was the 20s variable followed by the 60s, random, and 90s variables. These findings build empirical evidence for the use of a 60s, random, and 90s MTS procedure for the purpose of MVPA assessment by physical educators.

Accuracy of Momentary Time Sampling:
A Comparison of Varying Interval Lengths Using SOFIT

by

Jeff McNamee

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

Presented November 21, 2003
Commencement June 2004

Doctor of Philosophy dissertation of Jeff McNamee presented on November 21, 2003

Approved:

Redacted for Privacy

Major Professor, representing Health and Human Science

Redacted for Privacy

Chair of Department of Exercise and Sport Science

Redacted for Privacy

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

Redacted for Privacy

Jeff McNamee, Author

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

Accuracy of momentary time sampling:
A comparison of varying interval lengths using SOFIT

Introduction

The U.S. Department of Health and Human Services (USDHHS) has categorized physical inactivity as a major risk factor for several chronic diseases and has recently recognized it as equally detrimental to health as smoking (2000). Moderate levels of physical activity have been shown to enhance physical and mental health while producing positive changes in one's quality of life (Bouchard, Shepard, & Stephens 1994; Pate, Corbin, & Pangrazi, 1998; USDHHS, 1996). Investigations have demonstrated the significant association between moderate physical activity and the reduction of coronary heart disease, hypertension, colon cancer, and obesity (Blair, 1993; Freedson & Rowland, 1992; Kuntzleman & Reiff, 1992; Sallis & Patrick, 1994). Weight-bearing physical activities have been shown to increase bone mineral density (Fuchs, Bauer, & Snow, 2001), and vigorous exercise has been found to improve psychological health and mood (Calfas & Taylor, 1994). With these benefits in mind, the U.S. Public Health Service (2000) has made the promotion of regular physical activity a national health objective and one point is clear: Physical activity can be accumulated throughout the day. Specifically, experts recommend that youth accumulate 30-60 minutes of moderate physical activity daily (Pate, Corbin, & Pangrazi, 1998; Sallis & Patrick, 1994; USDHHS, 2000).

Physical education in schools is the only public institution currently responsible for promoting physical activity for all children (Sallis & McKenzie, 1991).

Experts believe physical education in schools is an investment in our society's economic and health care future (Corbin, 2002; Morrow, Jackson, & Payne 1999; Sallis & McKenzie, 1991). Professionals have outlined developing youth's physical activity behavior (i.e., process), rather than their physical fitness (i.e., product), as a major national health objective for physical education programs (USDHHS, 2000).

A shift in thinking from the Exercise Prescription Model (EPM), a model that focuses on measures of cardiovascular fitness (i.e., product), toward the Children's Lifetime Physical Activity Model (CLPAM), a model that focuses on the accumulation of moderate to vigorous physical activity throughout the day, is critical given that studies have demonstrated that children are the most physically active segment of our society and should not be treated like miniature adults (Simons-Morton et al., 1987). From a perspective of trying to reduce all-cause mortality, Simons-Morton et al. (1987) have argued that increasing youths' physical fitness should not be a major health priority. Freedson and Rowland (1992) agree, emphasizing that it would be more productive to focus on youths' physical activity levels, rather than fitness levels. Investigations have supported these suggestions as the relationship between youth's physical activity and their physical fitness is low (i.e., typical correlation of .16 - .17) (Morrow & Freedson, 1994).

As part of this paradigm shift from a product to a process orientation, individuals within the physical education and health professions have collaborated to emphasize what has been termed health-related physical education (HRPE), which promotes regular participation in physical activity (Corbin & Pangrazi, 1998; Pate &

Hohn, 1994; Sallis & McKenzie, 1991). An example of this paradigm shift can be witnessed in the Healthy People 2010 document published by the U.S. Public Health Service (2000). This document includes many health objectives that target physical activity behavior, including one that aims for students to engage in moderate to vigorous physical activity (MVPA) during at least 50% of the physical education class time. This recommendation is based on the premise that physical education should play a vital role in public health promotion (Sallis & McKenzie, 1991).

Morrow and Jackson (1999) stated, "it is important to develop an appreciation of physical activity and develop lifestyle behaviors in children and youth that can be adopted and maintained into adulthood" (p. 3). This approach attempts to assist individuals in establishing regular physical activity patterns that can be embraced as part of their everyday lives (Sallis & McKenzie, 1991). Increasing children's physical activity is paramount, given the evidence that health risk factors in childhood may predict health risk factors in adulthood (Cresanta et al., 1986). Coupled with the sharp decline in physical activity during the teenage years, Rowland (1999) has referred to adolescence as a risk factor for activity.

Although physical educators may embrace Healthy People 2010's MVPA goal, at present, there are limited ways for teachers to actually collect student MVPA data in a practical manner during class. Teachers could ask third parties to collect MVPA through systematic observation instruments or through heart rate monitoring. However, both of these methods would rely on "outside assistance" and do not allow the teacher to be directly involved in the process. If teachers are not directly involved

in the process while the lesson is occurring, adjustments to lesson based on MVPA data cannot occur.

Educational Reform and Authentic Assessment

Educational reform is another issue facing today's teachers, as they have been asked to find "authentic" ways to assess student learning on an on-going basis (Greenwood & Maheady, 1997, Morrow, Jackson, Disch, & Mood, 2000; NASPE, 1995; Wood, 1996; 2003). Wood (1996) asserts "physical education may be on the verge of an assessment revolution characterized by greater integration of assessment into the instructional process and development of alternative assessment strategies" (p. 213). *Authentic assessment* is an umbrella term for those strategies that broadly assess what students know and can do when they exit a grade level and tend to integrate higher order psychomotor, cognitive, and affective objectives under authentic conditions (Wood, 1996). Specific authentic assessment strategies include portfolios, discussions, debates, case studies, student activity logs, and role-playing. Siedentop and Tannehill (2000) have characterized authentic assessment as those events that reflect real life, are performed in realistic settings, and mirror what students do outside of school.

Resulting benefits of this assessment model are numerous as it "tends to engage students in the learning process; allows for the assessment of higher order cognitive processes; permits assessment of a combination of cognitive, psychomotor,

and affective behaviors; permits the assessment of what students know and can do in real-life contexts; and provides an assessment method for integrated learning across subject areas" (Wood, 1996, p. 214). It is believed that, through these processes, the authentic assessment model will provide a powerful means to link teaching, learning, and assessment (Wood, 1996).

Measurement of Physical Activity

Although a variety of other objective measures to assess human physical activity are available (e.g., heart rate monitors, motion sensors, and doubly weighted water), direct observation methods have several advantages over these techniques (McKenzie, Sallis, Nader, 1991; Dishman, Washburn & Schoeller, 2001) and have been utilized extensively in physical education teaching research (e.g., Keating, Kulinna, & Silverman, 1999; McKenzie, Sallis, & Nader, 1991; McKenzie et al., 1995; McKenzie, Marshall, Sallis, & Conway, 2000; Pope, Coleman, Gonzalez, Barron, & Heath, 2002; Rowe, Schuldheisz, & van der Mars, 1997; Schuldheisz & van der Mars, 2001). One such available objective measurement tool to directly observe students' physical activity patterns, System for Observing Fitness Instruction Time (SOFIT), has been validated for the purpose of collecting data on human physical activity behavior (Keating, Kulinna, & Silverman, 1999; McKenzie, Sallis, & Nader, 1991; Pope, Coleman, Gonzalez, Barron, & Heath, 2002; Rowe, Schuldheisz, & van der Mars, 1997). SOFIT provides a nonintrusive and inexpensive means of observing

6

and recording physical activity behaviors. SOFIT is recognized as a valid research tool in assessing associations of variations in both curriculum offerings and teaching behavior patterns (Rowe, Schuldheisz, & van der Mars, 1997; Schuldheisz & van der Mars, 2001; van der Mars, Vogler, Darst, & Cusimano, 1998).

Use and Validity of Momentary Time Sampling

Duration recording (DR) offers a complete record of the temporal dimension of behavior and serves as a measure to compare estimates of behavior (van der Mars, 1989). One such estimate technique, commonly used in applied behavior analysis, is called momentary time sampling (MTS) (van der Mars, 1989b). Unlike DR procedures, MTS provides an estimate of behavior that allows the user to take samples of behavior during an event (e.g., teaching episode) at the end of each observation interval (e.g., 10 seconds; van der Mars, 1989b). The measured targeted behavior is scored as having occurred if it was present at the end of the observation interval (see Figure 1).

One must be cautious when interpreting MTS data, as it only provides an estimate of true behavior (Test & Heward, 1984). When MTS procedures are compared to measures of the same behavior produced by continuous duration recording, it is possible that they can over- or underestimate the true behavior (Powell, Martindale, & Kulp, 1975). In general, underestimation may occur when the intervals are long and the typical duration of the target behavior is short. Overestimation may

Figure 1

Graphic Display of a 1 min. MTS Observation Tactic

Time elapsed (seconds)
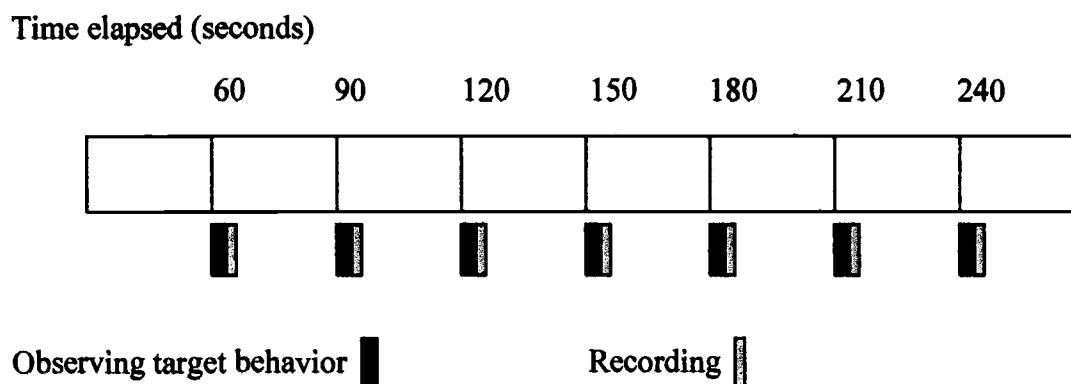
result when intervals are short and the typical duration of the target behavior is long.

When the frequency and length of intervals closely match, MTS procedures

offer excellent estimates of actual behavior when compared to DR (Leger, 1977;

Powell, Martindale, & Kulp, 1975; Powell, Martindale, Kulp, Martindale, Bauman,

1977; Simpson & Simpson, 1977). Generally, findings from these investigations have

demonstrated that a) MTS provides an accurate estimation of true behavior collected

through DR in lab settings and b) 120 seconds seems to be the maximal interval length

for MTS. Harrop and Daniels (1993) have encouraged, however, a "more

sophisticated mathematical treatment" (i.e., statistical techniques) of data collected

through MTS procedures as previous investigations have relied on visual analysis

alone.

Springer, Brown, and Duncan (1981) argued that MTS procedures highly distort records of true behavior as the recording of behavior depends upon the passage of time, rather than the true presence or absence of the behavior. However, investigations have shown that when time samples were conducted at intervals of 120 seconds or less, the time samples closely resembled the DR measure (Powell, Martindale, & Kulp, 1975).

Authentic Assessment of MVPA during Physical Education

The most unique advantage of MTS procedures compared to DR is that it allows practitioners to collect data *while* performing their regular duties (e.g. teaching) if the interval is long enough. Because teachers often times perform multiple tasks simultaneously during a teaching episode, they cannot collect data through continuous observation techniques.

It seems imperative that if physical educators are to embrace goals related to authentic assessment and physical activity behavior, they will need to be provided with practical assessment tools to gather valid data about their students' MVPA levels during physical education. Yet, instruments such as SOFIT have received no attention as a feasible tool to authentically assess on-going student MVPA during class by the physical educator. Additionally, the activity level section of the original SOFIT instrument utilizes a MTS collection procedure (10 seconds) that prevents teachers from simultaneously teaching and collecting valid in-class MVPA data for informal assessment purposes.

Therefore, this investigation sought to determine the accuracy of students' MVPA levels collected through MTS using varying interval lengths relative to those collected through duration recording. Specifically, the purpose of this investigation was to compare 20s, 60s, 90s, 120s, 180s, and random momentary time samples to duration recording.

Methods

Participants

Video records of 30 randomly selected elementary school physical education classes were obtained from a previously completed research project. Oregon State University's Institutional Review Board (IRB) approved this project for Human Subject research. Informed consent was obtained from 11 licensed physical education teachers in three surrounding school districts (see Table 1). The average years of teaching experience was 8.36, while the teachers' average age was 32.82. Five of the teachers were recognized for their outstanding teaching at either the state and/or national level.

After the 30 selected videotapes were identified for examination, MVPA data were collected on a randomly selected student within the range of the video lens. Six females and five males, ranging in age from 6 to 10, comprised the participant sample. The participant population was not restricted by gender, ethnic group, or disability. The decision to include 30 tapes was based on moderate effect size estimates for behavioral research (Park & Shutz, 1999; Thomas & Nelson, 2002).

Procedures

During the video records, teachers wore a wireless microphone to capture verbal interactions with students. The camera lens was kept on participating students

during the entire lesson. There were occurrences when the students were not visible through the lens due to surrounding physical obstacles.

Table 1

Teacher Demographics (N = 11)

	Female (n=6)		Male (n=5)	
	M	SD	M	SD
Average Years of Teaching Experience	10.17	6.49	6.20	4.76
Average Age	35.17	8.04	30.00	4.53

Participating teachers made no adjustments from their regular curricular content for the purposes of the prior investigation.

Program & Lesson Content

Participating teachers followed a four-part lesson format typically presenting introductory, fitness, skill, and game-related activities. During these portions of a lesson a multi-activity based curriculum, with an emphasis on fundamental motor skills (i.e., locomotor movement, throwing, and striking) was presented. A strong emphasis was placed on teacher directed movement education during introductory activities. The fitness portion of the lesson emphasized the major areas of health-

related fitness (i.e., cardiorespiratory, muscular strength/endurance, and flexibility). Specific health-related fitness activities included skipping, curl-ups, and stretching. Skill activities included rhythms, striking, catching, and tumbling. Generally, various tag games were used to close the lesson.

Data Collection

Prior research has demonstrated content- and concurrent-related validity of the original SOFIT instrument (Keating, Kulinna, & Silverman, 1999; McKenzie, Sallis, & Nader, 1991; Pope, Coleman, Gonzalez, Barron, & Heath, 2002; Rowe, Schuldheisz, & van der Mars, 1997). The original SOFIT instrument requires that an observer simultaneously assess physical activity, teacher behavior, and lesson context utilizing a combination of MTS and interval recording techniques (McKenzie et al., 1991). Physical activity levels are coded and quantified from an energy expenditure perspective. Additionally, SOFIT provides a section to assess teacher behavior (i.e., promotes fitness, demonstrates fitness, instructs generally, manages students or the environment, observes the class, and off-task behavior) and lesson context (i.e., general content, knowledge content, and motor content). However, only the physical activity portion of SOFIT was utilized for this investigation. The physical activity codes (i.e., Code 1 = lying down, Code 2 = sitting, Code 3 = standing, Code 4 = walking, Code 5 = very active) utilized in SOFIT have been correlated and validated using energy expenditure estimates, heart rates, and accelerometers (McKenzie et al., 1991; Rowe, Schuldheisz, & van der Mars, 1997).

Due to the multiple tasks physical educators must perform during a lesson, it may be difficult to assess students' MVPA utilizing the original SOFIT's coding protocol (i.e., Code 1 = lying down, Code 2 = sitting, Code 3 = standing, Code 4 = walking, Code 5 = very active). Recently, Rowe, van der Mars, Schuldheisz, and Fox (1997) validated SOFIT's codes for measuring physical activity of high school students, comparing corresponding heart rates with energy expenditure. High school-aged students completed a standardized protocol including lying, sitting, standing, walking, running, curl-ups and push-ups. The investigators used heart rates and energy expenditure (i.e., oxygen uptake) as criteria for concurrent validity. Three interesting findings from the Rowe, van der Mars, Schuldheisz, and Fox (1997) study, relevant to the current investigation, included: 1) Codes 1, 2, 3 were statistically different from Codes 4 and 5; 2) Codes 1, 2, and 3 were not statistically different from each other, and 3) Codes 4 and 5 were not statistically different from each other. Based on their findings, the authors suggested that it would be appropriate to merge the five original SOFIT Codes into two categories (i.e., Codes 1, 2, and 3 become "no MVPA and Codes 4 and 5 become "yes MVPA") (Rowe, van der Mars, Schuldheisz, & Fox 1997). Thus, for the purpose of this project the original SOFIT activity categories were merged into "no MVPA" and "yes MVPA" so that the observers needed only to make a dichotomous decision.

Intervals coded as "yes MVPA" were divided by the total intervals to obtain a percentage of intervals in MVPA for each targeted student during a particular interval condition (i.e., 20s, 60s, 90s, 120s, 180s). The unit of analysis for comparison against

DR was the average percentage of intervals in MVPA for each condition. Since DR offers a complete record of MVPA levels, it severed as the "gold standard" (see Table 2). Data were analyzed for each lesson component (i.e., introduction, fitness, skill, and game), first (i.e., introduction and fitness) and second half (i.e., skill and game), as well as for the total lesson.

Table 2

Target Behaviors and Observation Tactic

Target Behavior	Behavior Observation Tactic	Interval Length (seconds)	Unit of Analysis
MVPA[1]	Continuous Duration Recording	1	% of Time
MVPA[2]	MTS	20	% of Intervals
MVPA[2]	MTS	60	% of Intervals
MVPA[2]	MTS	90	% of Intervals
MVPA[2]	MTS	120	% of Intervals
MVPA[2]	MTS	180	% of Intervals
MVPA[3]	MTS	Random (60-90s)	% of Intervals

Note. [1] = Phase 1 of data collection; [2] = Phase 2 of data collection; [3] = Phase 3 of data collection

During Phase 1, MVPA data was collected throughout the entire lesson utilizing DR for the purposes of comparison against the varying interval lengths. A second-by-second decision regarding MVPA was made for each student so that the investigator could locate differences between the varying interval lengths and the DR condition. Microsoft® Excel03 was utilized to create a second by second archive of students' MVPA. Observers began collecting data as soon as the video chronometer began its count and continued collecting until it was turned off. Utilizing the second by second data, the investigator utilized the "ADD" and "COUNT" options in Microsoft® Excel03 to extract momentary time samples at 20, 60, 90, 120, 180, and random seconds (Phase 2). Lastly, the investigator utilized Pearson product-moment (PPM) correlation coefficients to analyze the DR and MTS data (Phase 3). Based on these initial findings the random numbers generation option in Microsoft® Excel03 was utilized to create 30 random interval lengths, one set of random intervals for each class. Based on the results of the PPM the acceptable range for this interval was between 60 and 90 seconds.

Management of Potential Observer Errors

Several considerations must be made in terms of observer error when collecting data through systematic observation techniques. The possibility of observer drift, observer expectancies, and observer reactivity should be minimized. When the observer begins to change coding rules and interpret behavior definitions differently,

(s) he is falling victim to observer drift. Causes may include long periods between instrument use, combining or mixing definitions from other instruments, and unclear behavior definitions (van der Mars, 1989b). Observer expectancies, or bias, may occur because the observer becomes aware of investigation's purpose. Finally, knowing that someone else will be checking the observer's reliability may produce observer reactivity. Observer reactivity can be reduced by telling the observer all observations will be screened, performing random inter-observer checks, and/or video-taping the sessions so that the observer and his or her assessor will never come in contact.

Observer Training and Reliability

The investigator was trained extensively (more than 50 hours) in the use of the SOFIT physical activity categories with pre-recorded "gold standard" videotapes of elementary and middle school physical education lessons. The training continued until a 90% or better inter-observer agreement (IOA) with an expert coder was reached, utilizing the Scored-Interval (S-I) method (van der Mars, 1989c).

To ensure observer reliability and guard against observer drift, IOA data were collected on targeted students' MVPA levels across 20% of randomly selected DR sessions. The S-I method, utilized for this study, is considered the most common and rigorous method to calculate IOAs (van der Mars, 1989c). The S-I method entails taking the observed agreements in each second interval and dividing it by the total number of observed agreements and disagreements and multiplying by 100

(van der Mars, 1989c). The investigator was unaware as to which classes were utilized for purposes of estimating reliability.

Data Analysis

In terms of research on human behavior, experts have defined validity as the extent to which observations scored by an observer match those of a predetermined standard for the same data (Safrit & Wood, 1989; Thomas & Nelson, 2002). More specifically, this type of validity has been termed concurrent validity in the measurement field (Safrit & Wood, 1989; Thomas & Nelson, 2002). It has been recommended that two pieces of evidence (i.e., relative and absolute accuracy) be demonstrated when investigating issues of concurrent validity (Wood, Maddalozzo, & Harter, 2002). Given these measurement standards, three analyses were used to demonstrate the relative and absolute accuracy of data sets for each of the varying interval lengths as compared to data collected with DR. These included, (a) Pearson product-moment (PPM) correlation coefficients to determine "relative accuracy", (b) repeated measures (RM) analysis of variance (ANOVA), and (c) average error (AE). The latter two serve as indicators of "absolute accuracy."

Relative Accuracy

Pearson product-moment correlation coefficients between each varying interval length and DR provided minimal evidence of concurrent validity and assisted in demonstrating relative accuracy.

Absolute Accuracy

Because the Pearson r analysis does not assess the magnitude of difference between the varying interval lengths and DR, trend analysis, utilizing repeated measures (RM) analysis of variance (ANOVA), and average error (AE), were also utilized. AE and RM ANOVA provided evidence of absolute accuracy at the individual level between the varying observation lengths and DR. Specifically, the RM ANOVA assisted in examining the trend differences (i.e., orthogonal contrasts) between each varying interval length and DR using the "linear" polynominal. The trend analysis option in RM ANOVA was utilized for this study because the researcher was interested in the error "trend" across varying interval lengths and DR rather than multiple comparisons among a repeated measure (e.g., trial 1, 2, and 3). The advantage of using the trend analysis is that the researcher need not worry about violating the assumption of sphericity (i.e., lack of correlations between repeated measures and demonstration of equal variance) and an inflated Type I error rate (Liu, 2002).

In contrast, AE is an estimate of the average amount that each varying interval length differed from the DR tactic. The AE statistic is often used to cross-validate prediction equations (see Jackson, 1989; Wood, 1989). Readers may be more familiar with names such as standard error or total error. AE was chosen for this study because

of its descriptive name. Because AE includes a comparison at the individual level in its computation, it was considered the strongest evidence of concurrent validity of the three analyses (Wood, Maddalozzo, & Harter, 2002).

Microsoft® Excel03 was used to compute AE for each varying MTS length $(sqrt[\Sigma(x-x')^2/(n-1)])$; where x represents the time the student was in MVPA calculated from the DR variable, x' represents the time spent in MVPA by the student under the varying MTS variable, and n equals the sample size. Given this study's exploratory nature, alpha was set at .05 (Thomas, Salazar, & Landers, 1991).

Results

The results are presented in five sections: interobserver agreements, mean

MVPA percentages, relative accuracy using PPM, trend statistics using RM ANOVA,

and AE for all measured MTS variables across 30 lessons. Each portion of the lesson

(i.e., introduction, fitness, skill, game, total, first half, and second half) had seven

accompanying variables (i.e., continuous duration recording and the varying interval

lengths) from which to collect data.

Interobserver Agreements

To ensure observer reliability and guard against observer drift, interobserver

agreement (IOA) data were collected on targeted students' MVPA levels across six

randomly selected duration recording (DR) sessions. The Scored-Interval

(S-I) method was utilized for this study and is considered the most rigorous method to

calculate reliability (van der Mars, 1989c). The range for the total class IOA's varied

from 76.00 to 100.00% with a mean of 88.71%. IOA data, partitioned into the four

lesson components, are listed in Table 3. Results of the IOA data suggest that the

investigator accurately coded each of the 30 tapes.

21

Table 3

Interobserver Agreement Percentages Across Lesson Components and Total Class

	Lesson Component				
Tape	Introduction (%)	Fitness (%)	Skill (%)	Game (%)	Total (%)
1	81.25	77.78	81.20	100.00	80.59
2	93.94	88.14	85.53	82.76	87.61
3	88.39	90.63	76.00	80.56	87.19
4	87.80	92.31	86.58	80.70	87.95
5	82.65	96.64	81.08	86.05	91.88
6	93.01	95.00	85.96	92.86	92.66

Mean MVPA Percentages

The average lesson length from which to collect MVPA was 27 minutes and 3 seconds (range of 22:08 - 30:00 min.). Systematic observation, utilizing a DR tactic, of students' MVPA levels over 30 physical education classes, varied from 11.61 - 32.92% with a mean percentage of 23.40%. MVPA levels were also calculated for the first half (i.e., introduction and fitness) and second half (i.e., skill and game) of the observed lessons. The MVPA for the first half of the lesson varied from 22.89 - 56.93% with a mean percentage of 35.65. The MVPA for the second half of the

lesson varied from 1.12% - 30.57% with mean percentage of 15.70. Categorized by the specific components of each lesson (i.e., introduction, fitness, skill, game), the MVPA data ranged from 7.42% - 74.36% with mean percentages equaling 34.93%, 35.55%, 15.29%, 16.77%, respectively. Descriptive statistics, including minimums, maximums, means, and standard deviations for each of the varying interval lengths are presented in Table 4, 5 and 6.

Relative Accuracy

Relative accuracy was examined by correlating the DR tactic with the varying lengths of MTS procedures using the Pearson product-moment (PPM) coefficient. A moderate to high positive correlation was expected for the 20s, 60s, and 90s MTS procedures (Harrop & Daniels, 1986; Leger, 1977; Powell, Martindale, & Kulp, 1975; Powell et al., 1977; Repp et al., 1976; Simpson & Simpson, 1977; Saudargas & Zanolli 1990). Correlations for the other MTS procedures were not known due to lack of prior research. The PPM coefficients for each part of the lesson as compared to DR tactic are presented in Table 7.

Correlation coefficients between the DR tactic and the MTS procedures, across the total lesson, ranged from .34 ($r^2 = .12$) to .74 ($r^2 = .55$). Specifically, correlation coefficients between the DR tactic and the varying MTS procedures (20s, 60s, 90s, 120s, 180s, random) equaled .74 ($r^2 = .55$), .69 ($r^2 = .48$), .42 ($r^2 = .18$), .54 ($r^2 = .29$), .34 ($r^2 = .12$), and .65 ($r^2 = .42$), respectively. The correlations between the DR tactic and the 20s, 60s, 120s, random MTS variables were significantly different from zero

at the p ≤ .01 level, while the 90s MTS variable was significant from zero at the p ≤ .05 level.

Table 4

Introduction and Fitness MVPA Descriptive Statistics for the Varying Interval Lengths

	Minimum	Maximum	M	SD
Introduction				
20	00.00	83.83	36.62	2.75
60	00.00	100.00	31.94	31.95
90	00.00	100.00	33.00	44.12
120	00.00	100.00	35.83	44.38
180	00.00	100.00	20.00	38.51
Ran	00.00	100.00	54.44	42.42
Fitness				
20	11.76	72.73	35.70	13.66
60	00.00	88.00	38.35	22.77
90	00.00	100.00	41.62	29.78
120	00.00	100.00	31.78	30.06
180	00.00	100.00	36.11	36.38
Ran	00.00	100.00	38.77	25.25

Note. M = mean; SD = standard deviation; Ran = random.

Table 5

Skill and Game MVPA Descriptive Statistics for the Varying Interval Lengths

	Minimum	Maximum	M	SD
Skill				
20	00.00	35.00	18.02	10.31
60	00.00	100.00	23.53	20.52
90	00.00	100.00	18.78	22.02
120	00.00	100.00	25.52	21.23
180	00.00	100.00	22.72	28.74
Ran	00.00	50.00	14.58	14.21
Game				
20	00.00	47.00	15.62	15.16
60	00.00	50.00	15.13	17.45
90	00.00	100.00	19.39	26.61
120	00.00	100.00	21.22	30.61
180	00.00	100.00	21.11	36.08
Ran	00.00	00.00	00.00	00.00

Note. M = mean; SD = standard deviation; Ran = random.

Table 6

First and Second Half of Class MVPA Descriptive Statistics for the Varying Interval Lengths

	Minimum	Maximum	M	SD
First Half				
20	8.70	58.33	36.09	11.14
60	14.29	71.43	38.36	17.31
90	00.00	100.00	42.63	22.35
120	00.00	83.33	35.33	27.04
180	00.00	100.00	41.49	29.72
Ran	00.00	77.78	40.37	20.54
Second Half				
20	00.00	30.43	17.15	8.72
60	00.00	44.44	19.85	12.39
90	00.00	50.00	17.78	14.85
120	00.00	50.00	21.89	14.50
180	00.00	75.00	20.35	19.61
Ran	00.00	45.45	14.65	12.69

Note. M = mean; SD = standard deviation; Ran = random.

The results from the PPM analysis indicate a moderate correlation between the DR tactic and 90s and 120s MTS procedure. Moderate to high correlations were found between the DR tactic and the 20s, 60s, and random MTS procedures.

Table 7

Correlations Between DR Tactic and the Varied MTS Interval Lengths for Each Lesson Component

MTS	Introduction DR	Fitness DR	Skill DR	Game DR
20	$.78^{**}$	$.86^{**}$	$.76^{**}$	$.85^{**}$
60	$.66^{**}$	$.58^{**}$	$.63^{**}$	$.41^{*}$
90	$.63^{**}$	$.38^{*}$	$.47^{**}$	$.48^{**}$
120	$.40^{*}$	$.36^{*}$	$.47^{**}$	$.39^{*}$
180	$.66^{**}$	$.49^{**}$	$.44^{*}$	$.40^{*}$
Random	$.44^{*}$	$.47^{*}$	$.57^{**}$	...

[*]Correlation significant from zero (p < .05, two-tailed).
[**]Correlation significant from zero (p < .01, two-tailed).

Trend Statistics Using RM ANOVA

Table 8 includes trend statistics over the full range of lesson context and MTS procedures. Type I error was set at $\alpha = .05$ (two-tailed). Therefore, $p \leq .05$ provided evidence of lack of trend similarity between the DR tactic and the varying MTS procedures. A RM ANOVA was utilized to examine significant differences between

the DR tactic and the MTS procedures. For models that demonstrated significance (i.e., lack of similarity), a Bonferroni alpha adjustment was used for the post hoc analysis. Nonsignificant differences were preferred to demonstrate similar trends in error between the DR tactic and the MTS observation tactics.

Table 8

Results of RM ANOVA Comparing DR with the Varied MTS Interval Lengths Across Lesson Components and Total Class

	Mean Square	F	Sig.	Partial Eta Squared	Observed Power
Intro	913.29	1.81	.189	.059	.255
Fitness	16.14	.054	.818	.002	.056
Skill	91.28	.605	.443	.020	.117
Game	1185.03	6.714	.015*	.188	.707
First Half	515.371	2.396	.132	.076	.322
Second Half	32.87	.377	.544	.013	.091
Total	20.95	.449	.508	.015	.099

Note. $df = 1$; *($p \leq .05$)

Results indicated no significant differences between DR and the varying interval lengths for the introduction, fitness, skill, first half, second half, and total portions of the lesson (see Table 8). The RM ANOVA, however, did detect lack of similarity between DR and the MTS procedures during the game portion of the lesson,

F(1, 29) = 6.71, p < .05; ES = .19.   Post hoc analysis showed that the mean difference

between the DR tactic and the random MTS procedure significantly contributed to this

difference, F(1, 29) = 27.64, p < .007); ES = .49.

Average Error (AE)

AE determines the difference between the DR tactic and the MTS tactics

across each class. The advantage of AE is that it accounts for the algebraic sign in its

computation by squaring the error, thus, providing the most accurate estimation of

error in comparing the DR tactic to the varying interval lengths.   Figure 2 displays the

magnitude of overestimation (+) and underestimation (-) of MVPA behavior across

interval lengths for each class over the total lesson.

Figure 2

Overestimation and Underestimation of MVPA Across Interval Lengths

MTS Over Total Lesson
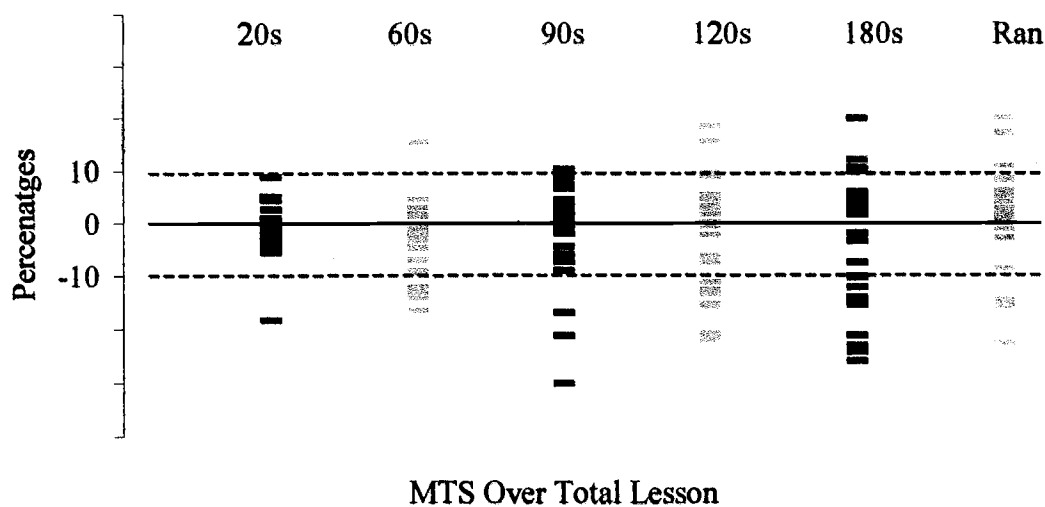
The "zero" line in Figure 1 represents MVPA behavior collected using the DR tactic

and represents the "true behavior." As shown in Figure 1, each MTS tactic created

measurement error and tended to both overestimate and underestimate the true MVPA

behavior. The MTS at 20s created the least amount of error across classes followed by

the 60s and 90s intervals. The MTS at 20s tended to equally over- and underestimate

true MVPA behavior. The MTS at 60s was more likely to underestimate MVPA

levels while at 90s it tended to overestimate this same behavior. The 120s, 180s, and

random interval tactics created the most measurement error and appeared to equally

over- and underestimate MVPA across classes.

To assist in examining the practical significance of AE, Table 9 provides AE

expressed as percentage of the mean difference between the DR tactic and the varying

interval lengths for the introduction, fitness, skill, and game portions of the lesson.

Table 10 provides the AE mean percentage difference for the first half, second half,

and total lesson.

AE of MVPA comparing the DR tactic with the varying interval lengths

ranged from $\pm$ 4.73% - $\pm$ 42.97%. Tables 9 and 10 show that the smallest AE across

the varying interval lengths was found between the DR tactic and the 20s MTS (range

of 4.73% - 14.48%). These tables also show the largest AE, across the varying

interval lengths, was found between the DR tactic and the 180s MTS procedure (range

of 12.40% - 33.36%). The average AE percentages across the varying interval lengths

(i.e., 20s, 60s, 90s, 120s, 180s, random) were 7.78, 15.82, 21.84, 24.46, 26.32, and

19.66. When examining AE, these data demonstrate that, on average, the 20s MTS

procedure was the most accurate, followed by the 60s, random, and then the 90s interval. When examining AE over the total lesson, Table 10 shows relatively small error differences between the DR tactic and the 20s ($\pm$5.63%), 60s ($\pm$10.09), random ($\pm$10.17), and the 90s ($\pm$13.47) MTS procedures.

Table 9

AE, Expressed as a Percentage, Between the DR Tactic and the Varying Interval Lengths across Lesson Components

MTS	Introduction DR	Fitness DR	Skill DR	Game DR
20	14.48	7.04	7.31	8.16
60	24.48	18.75	18.51	16.92
90	35.65	28.26	19.87	23.47
120	40.89	28.33	21.46	28.49
180	33.44	32.14	27.18	33.36
Random	42.97	22.86	11.71	21.26

Table 10

AE, Expressed as a Percentage, Between the DR Tactic and the Varying Interval
Lengths for the First, Second, and Total Lesson Components

MTS	First Half DR	Second Half DR	Total DR
20	4.73	7.10	5.63
60	7.07	14.90	10.09
90	9.92	22.25	13.47
120	11.24	26.50	14.33
180	12.40	27.61	18.08
Random	8.81	19.85	10.17

Discussion

The purpose of this investigation was to determine the accuracy of data on students' MVPA collected through MTS using varying interval lengths relative to those collected through DR. Specifically, the investigator compared MTS intervals of 20s, 60s, 90s, 120s, 180s, and random to DR. Because DR is considered to be a "true and complete record" of behavior it served as the "gold standard" against which the various MTS tactics were compared. The discussion section is presented in two subsections: descriptive data and MVPA data accuracy

Descriptive Data

MVPA data, collected through the DR tactic, of six female and five male elementary students revealed a mean MVPA of 23.40% (range of 11.61% - 32.92%) during the full course of a lesson. When MVPA data was examined for the first- and second half of the lesson, mean MVPA percentages equaled 35.65 and 15.70, respectively. We would expect the first half of the lesson to be substantially higher than the second half due to lesson content as it includes fitness related activities. Past investigations have demonstrated higher MVPA percentages during the introductory and fitness portions of the lesson as compared to the skill and game portions (McKenzie et al., 1995; McKenzie, Sallis, & Nader, 1991; Schuldheisz & van der Mars, 2001).

The average MVPA data, across the entire lesson, was slightly less than previous studies using SOFIT's MTS procedure (Levin, McKenzie, Hussey, Kelder, Lytle, 2001; McKenzie et al., 1995). Levin et al. (2001) investigated the magnitude and source of variability in physical activity during 324 indoor elementary school physical education lessons and reported MVPA means for third grade (29.4%), fourth grade (35.1%), and fifth grade (42.3%). When investigating children's physical activity levels and lesson context during third-grade physical education classes, McKenzie et al. (1995) reported a mean MVPA of 31.7%. Factors influencing these discrepancies may be attributed to lesson-to-lesson variability (Levin et al., 2001) and the presence of an intervention to increase students' MVPA (McKenzie et al., 1995; Schuldheisz & van der Mars, 2001). Additionally, these investigations utilized a 20s MTS procedure that may have produced measurement error.

The total class student MVPA data collected over 30 lessons, revealed these physical education classes were below the Healthy People 2010 goal of 50%. This is interesting as previous studies utilizing these teachers revealed that they were highly active in their supervision patterns, a teacher behavior found to be associated with higher levels of student MVPA (van der Mars, Vogler, Darst, Cusimano, 1998). This finding points to the importance of curricular selection by teachers as it seems the content implemented by these teachers, not their teaching behaviors, may have played a substantial role in producing relatively low MVPA levels. Of course, elementary physical education's goal is not entirely based on MVPA (NASPE, in press). Elementary teachers must find a balance toward meeting the Healthy 2010 MVPA

goal and ensuring that their students are well versed in fundamental skills, dance, and games.

MVPA Data Accuracy

With regards to the accuracy of the MTS procedures, readers must first determine their purpose for MVPA data collection. Readers should note that the following discussion is presented from both a clinical and research perspective. Readers are reminded that the overarching goal of this investigation was to determine whether physical educators could be offered a valid assessment tool that has been used previously as a research tool. Finally, it is difficult to compare past investigations testing the accuracy of MTS procedures as no study has demonstrated that these procedures are accurate from a statistical perspective.

Relative accuracy correlations obtained in this study demonstrated significantly higher correlations when the amount of time samples increased over the course of a lesson (e.g., 120s to 90s MTS). MTS intervals of 20s, 60s, and 90s produced correlations that fell within the acceptable range (i.e., moderate-low to high), during the full course of the lesson. The random interval, ranging from 60 - 90 seconds, showed promise as well with correlations in the moderate-low category. These findings demonstrate evidence for the accuracy and feasibility of 60s and 90s intervals over the full course of the lesson.

Trend statistics, using a RM ANOVA, also offered support for using the modified SOFIT activity level categories with longer intervals. RM ANOVA was

used to assess the mean differences between DR and varying interval lengths at the group level. Trend statistics showed no significant mean differences between DR and the varying interval lengths for the first half, second half, and total portions of the lesson. Lack of similarity was detected, however, between DR and the MTS procedures during the game portion of the lesson. This finding can be explained by a) the relatively short time span to collect MVPA and b) the lack of MVPA behavior during this portion of the lesson. Although the RM ANOVA demonstrates the accuracy of the MTS procedures during most components of the lesson, the utility of these statistics is limited by a) the failure to account for the reduction in overall mean error created by the canceling out effect of combining the over- and underestimates of the MTS procedures, b) low effect size estimates, and c) low statistical power.

Estimates produced from AE provide a clearer picture. AE was utilized to analyze the individual mean differences between DR and the varied interval lengths. Absolute differences were changed into a percentage of the mean difference between the DR tactic and the varying interval lengths across lesson context to assist in examining the practical significance of AE. When examining these percentages over the entire lesson, the 20s produced the least measurement error ($\pm$5.63%), then the 60s ($\pm$10.09), followed by the random ($\pm$10.17), and finally the 90s ($\pm$13.47) MTS procedure.

From a measurement perspective, it would appear the 60s, 90s, and random interval lengths provide sufficiently accurate estimates of elementary school-aged children's MVPA. These findings further support the use of a 20s MTS interval when

the intent is to collect the most accurate MVPA data for research purposes (Keating, Kulinna, & Silverman, 1999; McKenzie, Sallis, & Nader, 1991; Pope, Coleman, Gonzalez, Barron, & Heath, 2002; Rowe, Schuldheisz, & van der Mars, 1997).

Although the MTS accuracy results of this study are difficult to compare with past investigations due to differences in methodology, it appears this investigation produced similar results as those conducted by Harrop and Daniels (1986); Saudargas & Zanolli (1990); Powell, Martindale, & Kulp (1975); and Powell et al. (1977).

When Powell, Martindale, and Kulp (1975) analyzed the accuracy of time sampling recordings (i.e., 10s, 20s, 40s, 80s, 120s, 240s, 400s, and 600s) visually compared to DR the researchers found little difference (< 10%) between the 10s, 20s, 40s, 80s, and 120s time samples and DR measure. Similar to the present investigation when MTS procedures went beyond the 120s interval, larger discrepancies were observed. Powell et al. (1975) remind us that a summation of differences, or error, over sessions can "give a result that quite closely approaches a continuous measure" (p. 467). This is important to remember as the Healthy People 2010 MVPA goal implies that teachers would collect MVPA data over the full length of lesson. As evident by the results of all three analyses, when components of the lesson (e.g., introduction or game) only are utilized, more measurement error was produced. This increase in measurement error is due to the sensitivity of the MTS procedures over a very short time. Future investigations will need to examine the accuracy of MTS intervals (e.g., 120s) during longer class periods at the secondary level. Similar to the

conclusion made by Powell et al. (1975), MTS procedures became more accurate when a) the amount of the samples more closely approximates the true behavior and b) when the interval length closely matches the typical length of behavior.

The second point made by Powell, Martindale, and Kulp (1975) is critical when observing the MVPA behavior of elementary school-aged students and may explain why the accuracy of the 120-second interval was not replicated in the present study. Elementary students tend to participate in MVPA related activities for relatively short amounts of time with several periods of rest. If the MTS procedure goes beyond these short moments of MVPA behavior, over short amounts of lesson time (e.g., introduction), it is more likely that these observation tactics will not be sensitive enough to capture MVPA behavior accurately.

The sensitivity of MTS procedures was apparent in Harrop and Daniels' (1986) investigation when these authors compared MTS data against measures of behavior from DR. The authors indicated the that sensitivity of time sampling methods increases with lower rates of responding and longer emitted duration of the target behavior. Thus, it may well be that a 90s MTS interval is the longest possible length for measuring MVPA behavior of elementary students. Future investigations will be needed to test more precise time intervals between 90s and 120s (e.g., 95s, 100s, 105s, etc) to demonstrate whether this is the case. Additionally, future investigations will need to demonstrate evidence of validity for MTS procedures of secondary students, as their natural physical activity patterns are different from that of elementary students.

From a clinical perspective, these data offer important information in support of utilizing MTS tactics for assessment of MVPA by teachers. With regard to the act of teaching and on-going formative assessment, the results of this investigation have four practical implications for physical educators. First, teachers can use MTS tactics with interval lengths up to 90 seconds. Second, teachers can choose between a 60s, 90s, and random interval length. Third, they offer teachers a way to collect accurate MVPA data on a physical education program's ability to promote physical activity during class. Lastly, the integration of assessment with instruction can, and should, be considered.

Conclusions

The major findings of this investigation were as follows:

1. When compared to DR, MTS procedures produced measurement error across lesson context.

2. Moderate-low to high correlations were found between DR and 20s, 60s, and 90s MTS procedures.

3. Similarity statistics showed that there were no significant mean differences between the DR and MTS procedures over the full course of the lesson.

4. Both absolute and relative AE were relatively low ($\leq 10\%$) for the 20s, 60s, and random MTS procedures across the entire lesson. The AE for the 90s interval length was $\pm 13.47\%$.

5. The accuracy of the MTS procedures examined in this study varied over different lesson contexts.

6. The utilization of the 20s MTS length was validated from a research perspective.

7. The 60s, 90s, and random MTS lengths showed the most promise from a clinical perspective.

This study found that when MTS procedures are compared against DR, measurement error increases as the number of samples taken per lesson decreases (i.e., the MTS interval is stretched further). The strength of a 20s MTS was replicated in this investigation and is considered to be a strong research tool to estimate MVPA behavior in elementary physical education class settings. From a clinical perspective, concurrent-related evidence of validity for the 60s, 90s, and random interval lengths over the course of a full lesson was established. Readers will need to determine the

acceptable amount of measurement error they are willing to accept based on their purpose for MVPA assessment, age level of students, and program context.

The results of this study indicate the accuracy and feasibility of systematic observation as a tool to collect on-going formative MVPA behavior of students in physical activity settings. As noted by Test and Heward (1984), the major advantage of MTS procedures is that teachers can utilize them *during* an instructional episode. It is further hypothesized that the task of MVPA data collection will be alleviated as the length of the MTS lengthens. For this reason, this project found that teachers could utilize a 60s or 90s MTS length to collect on-going formative MVPA data. The researchers are not suggesting that MVPA data be used in a summative manner for grades or other formal assessments. Rather, MVPA data might be collected to assist teachers in understanding the relationship between their teaching behaviors and choice of lesson content with student MVPA behavior. MVPA data could also be utilized to assist teachers in demonstrating their ability to reach the Healthy People 2010 MVPA goal. From a student perspective, MVPA data could be utilized to assist students in understanding the amount of physical effort they are putting forth during a lesson and linking this behavior to their overall physical health and fitness.

Future investigations, in different physical activity settings, school contexts, and varying ages of students will be needed to further establish the generalizability of these findings. Additionally, future investigations will be needed to demonstrate that teachers themselves can collect reliable data during the act of teaching.

References

Blair, S.N. (1993). 1993 C.H. McCloy research lecture: Physical activity, physical activity, and health. *Research Quarterly for Exercise and Sport, 64,* 365-376.

Blair, S.N., Kohl, H.W., & Gordon, N.F. (1992). Physical activity and health: A lifestyle approach. *Medicine, Exercise, Nutrition, and Health, 1,* 54-57.

Bouchard, C., Shepard, R., & Stephens, T. (1994). *Physical activity, fitness, and health: International proceedings and consensus statement.* Champiagn, IL: Human Kinetics.

Bouchard, C., & Despres, J.P. (1995). Physical activity and health: Atherosclerotic, metabolic, and hypertensive diseases. *Research Quarterly for Exercise and Sport 66* (4), 268-276.

Brands, P. (1996). The pros outweigh the cons: An authentic assessment approach to teaching physical education. *Teaching Secondary Physical Education, 2* (4), 18-21.

Brosse, A.L., Sheets, E.S., Lett, H.S., & Blumenthal, J.A. (2002). Exercise and the treatment of clinical depression in adults: recent findings and future directions. *Sports medicine 32* (12), 741-760.

Calfas, K.J., & Taylor, W.C. (1994). Physical activity and psychological health in adolescents. *Pediatric Exercise Science, 6,* 406-423.

Caspersen, C.J., Powell, K.E., Christenson, G.M. (1985). Physical activity, exercise and physical fitness: definitions and distinctions for health-related research. *Public health reports, 100* (2), 126-131.

Cleland, F.E., & Stevenson, R. (1997). Authentic assessment in action. *Teaching Secondary Physical Education, 3* (3), 4-7.

Corbin, C. (2002). Physical activity for everyone: What every physical educator should know about promoting lifelong physical activity. *Journal of Teaching in Physical Education, 21* (2), 128-144.

Corbin, C., LeMasurier, G, & Franks, B.D. (2002). Making sense of multiple physical activity recommendations. *President's Council on Physical Fitness and Sports Research Digest, 3* (19), 1-8. Washington, DC: The President's Council on Physical Fitness and Sports.

Corbin, C, & Pangrazi, B. (1996). How much physical activity is enough? *Journal of Physical Education, Recreation, and Dance, 64* (4), 33-37.

Corbin, C, & Pangrazi, B. (1998). *Physical activity for children: A statement of guidelines.* Reston, VA: National Association of Sport and Physical Education.

Corbin, C., & Pangrazi, B. (1998). Physical activity and aging: Implications for health and quality of life in older adults. *President's Council on Physical Fitness and Sports Research Digest 3* (4). Washington, DC: The President's Council on Physical Fitness and Sports.

Corbin, C.B., Pangrazi, R.P., & Welk, G.J. (1994). Toward an understanding of appropriate physical activity levels for youth. *Physical activity and Fitness Research Digest, 1* (8), 1-8. Washington, DC: The President's Council on Physical Fitness and Sports.

Cresenta, J.L., Burke, G.L., Downey, A.M., Freedman, D.S., & Berenson, G.S. (1986). Prevention of arteriosclerosis in childhood: Prevention in primary care. *Pediatric Clinics of North America, 33,* 835-858.

Dishman, R.K., Washburn, R.A., Schoeller, D.A. (2001). Measurement of physical activity. *Quest, 53,* 295-309.

Freedson, P.S., & Rowland, T.W. (1992).Youth activity versus youth fitness: Let's direct our efforts. *Research Quarterly for Exercise and Sport, 63,* 133-136.

Fuchs, R.K., Bauer, J.J., & Snow, C.M. (2001). Jumping improves hip and spine bone mass in pre-adolescent boys and girls: A randomized controlled trial. *Journal of Bone and Mineral Research, 16* (1), 148-156.

Greenwood, C.R., & Maheady, L. (1997). Measurable change in student performance: Forgotten standard in teacher preparation. *Teacher Education and Special Education, 20* (3), 265-275.

Harrop, A., & Daniels, M. (1986). Methods of time sampling: A reappraisal of momentary time sampling and partial interval recording. Journal of Applied Behavior Analysis, 19, 73-77.

Harrop, A., & Daniels, M. (1993). Further reappraisal of momentary time sampling and partial interval recording. Journal of Applied Behavior Analysis, 26, 277-278.

Haskell, W.L. (1985). Physical activity and health: Need to define the required stimulus. *American Journal of Cardiology, 5,* 4D–9D.

Jackson, A.S. (1989). Application of regression analysis to exercise science. In M.J. Safrit & T.M. Wood (Eds.), *Measurement concepts in physical education and exercise science* (pp. 181-205). Champaign, IL: Human Kinetics.

Karvonen, M.J. (1959). The effects of vigorous exercise on the heart. In Rosenbaum, F.F., & Belknap, E.L. (Eds.), *Work and the heart.* New York: P.B. Hoebner.

Keating, X.D., Kulinna, P.H., & Silverman, S. (1999). Measuring teaching behaviors, lesson context, and physical activity in school physical education programs: comparing the SOFIT and the C-SOFIT instruments. *Measurement in Physical Education and Exercise Science, 3* (4), 207-220.

Kuntzleman, C.T., & Reiff, G.G. (1992). The decline in American children's fitness levels. *Research Quarterly for Exercise and Sport, 63* (2), 107-111.

Landers, D.M. (1997). The influence of exercise on mental health. *Physical activity and Fitness Research Digest, 2* (12), 1-8. Washington, DC: The President's Council on Physical Fitness and Sports.

Lee, I.M., Paffenbarger Jr., R.S. (2001). Preventing coronary heart disease: The role of physical activity. *Physician and Sportsmedicine, 29* (2), 37-40; 43-46; 49; 52.

Lee, I.M., Rexrode, K.M., Cook, N.R., Manson, J.E., & Burning, J.E. (2001). Physical activity and coronary heart disease in women: Is "no pain, no gain passé? *Journal of the American Medical Association, 285* (1), 1447-1454.

Leger, D.W. (1977). An empirical evaluation of instantaneous and one-zero sampling in chimpanzee behavior. *Primates, 18,* 387-392.

Liu, Y. (2002). Analyzing RM ANOVA related data using SPSS10. *Measurement in Physical Education and Exercise Science, 6* (1), 43-60.

Levin, S., McKenzie, T.L., Hussey, J.R., Kelder, S.H., & Lytle, L.A. (2001). Variability of physical activity during education lessons across elementary school grades. *Measurement in Physical Education and Exercise Science, 5* (4), 208-218.

Manson et al. (2002). Walking compared with vigorous exercise for the prevention of cardiovascular events in women. *New England Journal of Medicine, 347* (10), 716-725.

Manson et al. (1999). A prospective study of walking as compared with vigorous exercise in the prevention of coronary heart disease in women. *New England Journal of Medicine, 341* (9), 650-658.

McKenzie, T.L., & Sallis, J.F. (1996). Physical activity, fitness, and health-related physical education. In S.J. Silverman & C.D. Ennis (Eds.), *Student learning in physical education: Applying research to enhance instruction* (pp. 223-246). Champaign, IL: Human Kinetics.

McKenzie, T.L., Sallis, J.F., Nader, P.R., Patterson, T.L., Elder, J.P., Berry, C.C., et al. (1991). BEACHES: An observational system for assessing children's eating and physical activity behaviors and associated events. *Journal of Applied Behavior Analysis, 24* (1), 141-151.

McKenzie, T.L., Feldman, H., Woods, S.E., Romero, K.A., Dahlstrom, V., Stone, E.J. et al. (1995). Children's activity levels and lesson context during third-grade physical education. *Research Quarterly for Exercise and Sport, 66,* 184-193.

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

McKenzie, T.L., Marshall, S.J., Sallis, J.F., & Conway, T.L. (2000). Student activity levels, lesson context, and teacher behavior during middle school physical education. *Research Quarterly for Exercise and Sport, 71,* 249-259.

Melograno, V.J. (2000). Designing a portfolio system for k-12 physical education: A step-by-step. *Measurement in Physical Education and Exercise Science, 4* (2), 97-115.

Morrow, J.R., & Freedson, P.S. (1994). Relationship between habitual physical activity and aerobic fitness in adolescents. *Pediatric Exercise Science, 6,* 315-329.

Morrow, J.R., Jackson, A.W., & Payne, V.G. (1999). Physical activity promotion and school physical education. In C.B. Corbin & R.P. Pangrazi (Eds.), *Physical Activity and Fitness Digest, 3* (7). Bloomington, IN: Indiana University.

National Association for Sport and Physical Education. (in press). *Physical activity and children: A statement of guidelines 2003.* Reston, VA: NASPE.

National Association for Sport and Physical Education. (1995*). Moving into the future: National standards for physical education.* Reston, VA: AAHPERD.

Nichols, J.F., Morgan, C.G., Chabot, L.E., Sallis, J.F., Calfas, K.J. (2000). Assessment of physical activity with the computer science and applications, inc., accelerometer: Laboratory versus field validation. *Research Quarterly for Exercise and Sport, 71*, 36-43.

O'Hara, N., Baranowski, T., Simons-Morton, B., Wilson, S., & Parcel, G. (1989). Validity of the observation of children's physical activity. *Research Quarterly for Exercise and Sport, 60*, 42-47.

Paffenbarger, R.S., Hyde, R.T., Wing, A.L., & Hsueh, R.T. (1986). Physical activity, allcause mortality, and longevity of college alumni. *New England Journal of Medicine, 314*, 605–613.

Paffenbarger, R.S., Hyde, P.T., Wing, A.L., & Steenmetz, C.H. (1984). A natural history of athleticism and cardiovascular health. *Journal of the American Medical Association, 252*, 491–495.

Paffenbarger, R.S., Wing, A.L., & Hyde, R.T. (1978). Physical activity as an index of heart attack risk in college alumni. *American Journal of Epidemiology, 108*, 161–175.

Pate, R., Corbin, C., Pangrazi, B. (1998). Physical activity for young people. *President's Council on Physical Fitness and Sports research digest 3* (3). Washington, DC: The President's Council on Physical Fitness and Sports.

Pate. R.R. & Hohn, R.C. (Eds.). (1994). *Health and fitness through physical education.* Champaign, IL: Human Kinetics.

Pate, R.R., Dowda, M., & Ross, J.R. (1991). Associations between physical activity and physical fitness in American children. *American Journal of Diseases of Children, 144*, 1123-1129.

Pope, R.P., Coleman, K.J., Gonzalez, E.C., Barron, F., & Heath, E.M. (2002). Validity of a revised System for Observing Fitness Instruction Time. (SOFIT). *Pediatric Exercise Science, 14* (2), 35-146.

Powell, J., Martindale, A., Kulp, S., Martindale, A., Bauman, R. (1977). Taking a closer look: Time sampling and measurement error. *Journal of Applied Behavior Analysis, 10*, 325-332.

Powell, J., Martindale, A., Kulp, S. (1975). An evaluation of time-sample measures of behavior. *Journal of Applied Behavior Analysis, 8*, 463-469.

Repp, A.C., Roberts, D.M., Slack, D.J., Repp, C.F., & Berkler, M.S. (1976). A compison of frequency, interval, and time-sampling methods of data collection. *Journal of Applied Behavior Analysis, 9,* 501-508.

Richard, J.F., Godbout, P., & Griffin, L.L. (2002). Assessing game performance: An introduction to the Team Sport Assessment Procedure (TSAP). *Physical & Health Education Journal, 68* (1), 12-18.

Ross, J.G., & Gilbert, G.G. (1985). The national children and youth fitness study: A summary of findings. *Journal of Physical Education, Recreation, and Dance, 56* (1), 45-50.

Ross, J.G., & Pate, R.R. (1987). The national children and youth fitness study II: A summary of findings. *Journal of Physical Education, Recreation, and Dance, 58* (9), 51-56.

Rowe, P.J., Schuldheisz, J.M., & van der Mars, H. (1997). Measuring physical activity in physical education: Validation of the SOFIT direct observation instrument for use with first to eighth grade students. *Pediatric Exercise Science, 9,* 136-149.

Rowe, P., van der Mars, H., Schuldheisz, J.M., & Fox, S. (1997). Measuring physical activity: Validating SOFIT for use with high school students. Paper presented at the 4th. Pacific Rim Conference on Exercise Science and Sports Medicine, Corvallis, OR.

Rowland, T.W. (1999). Adolescence: A "risk factor" for physical inactivity. *The President's Council on Physical Fitness and Sports Research Digest, 3* (6). Washington, DC: The President's Council on Physical Fitness and Sports.

Sallis, J.F., & McKenzie, T.L. (1991). Physical education's role in public health. *Research Quarterly for Exercise and Sport, 62* (2), 124-137.

Sallis, J.F., & Patrick, K. (1994). Physical activity guidelines for adolescents: Consensus statement. *Pediatric Exercise Science, 6,* 302-314.

Saudargas, R.A., & Zanolli, K. (1990). Momentary time sampling as an estimate of percentage time: A field validation. *Journal of Applied Behavior Analysis, 23,* 533-537.

Schuldheisz, J.M., & van der Mars, H. (2001). Active supervision and students' physical activity in middle school physical education. *Journal of Teaching in Physical Education, 21* (1), 75-90.

Siedentop, D. (1994). Authentic assessment through sport education. In, Siedentop, D. (Ed.), *Sport education: Quality PE through positive sport experiences* (pp. 115-118). Champaign, IL: Human Kinetics.

Siedentop, D., & Tannehill, D. (2000). *Developing teaching skills in physical Education (4th ed.)*. Mountain View, CA: Mayfield Publishing.

Simon-Morton, B.G., O'Hara, N.M., Parcel, G.S., Huang, I.W., Baranowski, T., & Wilson, B. (1990). Children's frequency of participation in moderate to vigorous physical activities. *Research Quarterly for Exercise and Sport, 61,* 307-314.

Simpson, M.J., & Simpson, A.E. (1977). One-zero and scan methods for sampling behavior. *Animal behavior, 25,* 726-731.

Smith, T.K., & Cestaro, N.G. (1981). Student-centered physical education: Strategies for developing middle school fitness and skills. Champaign, IL: Human Kinetics.

Sofranko, A.J., & Nolan, M.J. (1972). Early life experiences and adult sports participation. *Journal of Leisure Research, 9,* 6-18.

Springer, B, Brown, T., & Duncan, P.K. (1981). Current measurement in applied behavior analysis. *The Behavior Analyst, 4* (1), 19-31.

Strath, S.J., Swartz, A.M., Bassett Jr., D.R., O'Brien, W.L., King, G.A., Ainsworth, B.E. (2000). Evaluation of heart rate as a method for assessing moderate intensity physical activity. *Medicine and Science in Sports and Exercise, 32* (Suppl. 9), S465-S470.

Tanasescu et al. (2002). Exercise type and intensity in relation to coronary heart disease in men. *Journal of the American Medical Association, 288* (16),1994-2000.

Test, D.W., & Heward, W.L. (1984). Accuracy of momentary time sampling: A comparison of fixed- and variable-interval observation schedules. In W.L. Heward, T.E., Heron, D.S. Hill, & J. Trap-Porter (Eds.), *Focus on behavior analysis in education* (pp. 177-194). Columbus, OH: Merrill.

Thomas, J.R., & Nelson, J.K. (2001). *Research methods in physical activity.* Champaign, IL: Human Kinetics.

U.S. Department of Health and Human Services. (1996). *Physical activity and health: A report of the surgeon general.* Atlanta, GA: U.S. Department of HHS.

U.S. Department of Health and Human Services. (2000). *Healthy people 2010.* Washington, DC: U.S. Gov. Printing Office.

van der Mars, H. (1989a). Systematic observation: An introduction. In P. Darst, D. Zakrajsek, & V. Mancini (Eds.), *Analyzing physical education and sport instruction* (pp. 3-17). Champaign, IL: Human Kinetics.

van der Mars, H. (1989b). Basic recording techniques. In P. Darst, D. Zakrajsek, & V. Mancini (Eds.), *Analyzing physical education and sport instruction* (pp. 19-51). Champaign, IL: Human Kinetics.

van der Mars, H. (1989c). Observer reliability: Issues and procedures. In P. Darst, D. Zakrajsek, & V. Mancini (Eds.), *Analyzing physical education and sport instruction* (pp. 53-93). Champaign, IL: Human Kinetics.

van der Mars, H., Vogler, B., Darst, P., & Cusimano, B. (1998). Students' physical activity levels and teachers' active supervision during fitness instruction. *Journal of Teaching in Physical Education, 18* (1), 57-75.

Veal, M.L. (1992). The role of assessment in secondary physical education: A pedagogical view. *Journal of Physical Education, Recreation, and Dance, 63* (7), 88-92.

Weinberg, H. (1996). Authentic assessment in middle school physical education. *Teaching Secondary Physical Education, 2* (2), 20-21.

Welk, G.J., Differding, J.A., Thompson, R.W., Blair, S.N., Dziura, J., Hart, P. (2000). The utility of the Digi-Walker step counter to assess daily physical activity patterns. *Medicine and Science in Sports and Exercise, 32* (Suppl. 9), S481-S488

Wood, T.M. (2003). In S.J. Silverman & C.D. Ennis (Eds.), *Student learning in physical education: Applying research to enhance instruction* (pp. 199-219). Champaign, IL: Human Kinetics.

Wood, T.M. (1989). The changing nature of norm-referenced validity. In M.J. Safrit & T.M. Wood (Eds.), *Measurement concepts in physical education and exercise science* (pp. 23-44). Champaign, IL: Human Kinetics.

Wood, T.M (1996). Evaluation and testing: The road less traveled. In S.J. Silverman & C.D. Ennis (Eds.), *Student learning in physical education: Applying research to enhance instruction* (pp. 199-219). Champaign, IL: Human Kinetics.

Wood, T.M., Maddalozzo, G.F., & Harter, R.A. (2002). Accuracy of seven equations for predicting 1-RM performance of apparently healthy, sedentary older adults. *Measurement in Physical Education and Exercise Science, 6* (2), 67-94.

Appendices

Appendix A

Review of the Literature

The purpose of this review of the literature is to provide a synopsis of published work in the areas of a) the benefits of physical activity, b) the development of youth physical behavior, c) the role physical education in schools can play in promoting physical activity, d) methods of physical activity measurement, e) systematic observation and the use of momentary time sampling, f) the System for Observing Fitness Time Instruction, g) and a rationale to find authentic methods for physical activity assessment during physical education.

Health Benefits of Physical Activity

Before discussing the health benefits of physical activity it may be necessary to review a few terms predominant in the current literature. Sallis and Patrick (1994) have noted, the terms *physical activity*, *exercise*, and *physical fitness* are sometimes used synonymously. However, this is incorrect as each term is a distinct concept with different goals and outcomes.

The widely used definition of physical activity, a broader term then either exercise or fitness, has been described as "any bodily movement produced by skeletal muscle that results in energy expenditure" (Caspersen, Powell, & Christenson, 1985, p. 126). Because "results in energy expenditure" is a component of the definition, activities occupational in nature, household chores, leisure, sport participation, and activity that is planned for fitness are included in its definition. The term exercise

sometimes brings looks of trepidation on people's faces as some have equated this term with a "no pain, no gain" attitude. A subset of physical activity, exercise is "planned, structured, and repetitive bodily movement done to improve or maintain one or more components of physical fitness (Caspersen, Powell, & Christenson, 1985, p. 126). A popular term often misused in the media, physical fitness, is "a set of attributes that people have or achieve that relates to the ability to perform physical activity" (Caspersen, Powell, & Christenson, 1985, p. 126). Physical fitness can be separated into two components, *sport-related physical fitness* and *health-related physical fitness*. Sport-related physical fitness includes a) static and dynamic balance, b) power, c) agility, d) speed, and e) coordination (Pangrazi, 2001). Although these components are critical from a sports performance perspective they are not key indicators of physical health and are unrelated to reducing all-cause mortality. Components such as cardiorespiratory endurance, muscular endurance, muscular strength, flexibility, and body composition help to comprise health-related physical fitness. Health-related physical fitness components contribute to sports performance, the capacity to perform occupational tasks, and are directly related to physical health (Sallis & Patrick, 1994).

The benefits of regular participation in moderate physical activity is well documented in the literature and the U.S. Department of Health and Human Services (USDHHS) has issued physical activity statements and guidelines for youth,

adolescents, as well as, adults (2000). The Office of the Surgeon General has also recognized physical *inactivity* as a major risk for cardiovascular disease and has developed national health objectives to promote regular amounts of physical activity (USDHHS, 2000). Presently, research results are still accumulating in support of regular amounts of moderate and vigorous physical activity irrespective of race, age, and body mass, in terms of preventing all-cause morbidity and mortality (Blair, Kohl, & Gordon, 1992; Bouchard & Despres, 1995; Lee, Rexrode, Cook, Manson, & Burning, 2001; Manson et al., 2002; Manson et al., 1999; Tanasescu et al., 2002)

Benefits derived from regular amounts of physical activity reach much further then the reduction of cardiovascular disease, as it has also been associated with reducing the risks of several other significant chronic diseases. These include: a) increased immune functioning and longevity, b) reduction in adult onset diabetes, obesity hypertension, osteoporosis, and some types of cancer (Blair, 1993; Freedson & Rowland, 1992; Kuntzleman & Reiff, 1992; Lee & Paffenbarger, 2001; Paffenbarger, Hyde, Wing, & Hsueh, 1986; Paffenbarger, Wing, & Steenmetz, 1984; Paffenbarger, Wing, & Hyde, 1978; Rowland, 1991; Sallis & Patrick, 1994).

Investigations linking physical activity and mental health are becoming more predominant in the literature at this time preliminary studies are showing promise (Brosse, Sheets, Lett, & Blumenthal, 2002; Calfas & Taylor, 1994; Landers, 1997). Studies have found physical activity can have a positive impact on those individuals suffering from feelings of depression (Brosse et al., 2002; Landers, 1997) as well as

significantly lowering feelings of anger and stress in sedentary individuals (Calfas &

Taylor, 1994; Hassmen, Koivula, & Uutela, 2000).

Development of Youth Physical Activity

In 1996, the Surgeon General issued a report pertaining to the physical activity

behaviors of the general population (USDHHS, 1996). The report outlined several

benefits of physical activity and specifically called for the reduction of sedentary

living among the general population. According to Corbin, LeMasurier, and Franks

(2002) the Surgeon General's Report was significant for two reasons. First, a

paradigm shift from physical fitness to the promotion of the accumulation of physical

activity was apparent. Second, the report was a collaborative project created by

several governmental agencies, as well as professional and private organizations.

Based on data from the National Children and Youth Fitness Study (Ross &

Gilbert, 1985) evidence that health risk factors in childhood may predict health risk

factors in adulthood (Cresanta et al., 1986), the Surgeon General stated in his report

that school-based physical education programs are our most vital resource for physical

activity promotion (USDHHS, 1996). Sallis and McKenzie (1991) agreed and noted

that physical education in schools is the primary public institution currently

responsible for promoting physical activity for all children. Experts believe physical

education in schools is an investment in our society's economic and health care future

(Corbin, 2002; Morrow, Jackson, & Payne 1999; Sallis & McKenzie, 1991).

Professionals have outlined developing youth's physical activity behavior (i.e.,

process), rather than their physical fitness (i.e., product), as a major national health

objective for physical education programs (Corbin, 2002; Morrow, Jackson, & Payne

1999; Sallis & McKenzie, 1991; USDHHS, 2000).

The Council on Physical Education for Children, a subgroup of the National

Association for Sport and Physical Education (NASPE, 1998), developed detailed

physical activity guidelines for youth. Past guidelines suggest that children should

participate in physical activities that are 10 to 15 minutes in length, with brief periods

of rest and recovery (Corbin & Pangrazi, 1998). These guidelines have more recently

been revised (NASPE, in press) and state that children should be physically active at

least 60 minutes and up to several hours per day. Individuals working with children

should not ask them to participate in long bouts of vigorous activity, as children tend

to participate in vigorous to moderate activities intermittently with several periods of

rest. Sallis & McKenzie (1991) suggest moderate activities for youth, with numerous

periods of rest, as this regime is more likely to be maintained throughout adulthood.

Experts believe developing youth physical activity behavior should be the

responsibility of both public health and school professionals (Sallis & McKenzie,

1991). Sallis & McKenzie (1991) stated that elementary schools should emphasize

health related physical activities that will carry over in adulthood. The assumption,

supported by some research (Cresanta et al., 1986), is that physically active children

tend to be physically active adults and, thus, reduce their risk of cardiovascular and other chronic diseases (Sallis & McKenzie, 1991).

Experts have made a clear distinction between two approaches toward physical activity and health for adults: The Exercise Prescription Model (EPM) and the Lifetime Physical Activity Model (LPAM) (Corbin, Pangrazi, & Welk, 1994; McKenzie & Sallis, 1996). The EPM was formed in the early 1960s and focuses on higher intensity, shorter duration, activity using percentage of maximum heart as the criterion for intensity (Karvonen, 1959). Central to the EPM are measures of cardiovascular fitness (i.e., product).

The LPAM, however, focuses on accumulating 30 minutes of moderate to vigorous physical activity (MVPA) throughout the day on most days of the week (i.e., process; Haskell, 1994). Strong evidence exists to support the LPAM and results point to the numerous health benefits of lower intensity, longer duration, physical activity (Paffenbarger, Hyde, Wing, & Hsueh, 1986; Paffenbarger, Hyde, Wing, & Steenmetz, 1984; Paffenbarger, Wing, & Hyde, 1978). Investigators have found that common lifestyle activities such as climbing stairs, walking, and doing physically active household chores can significantly reduce the risk of chronic disease (Paffenbarger, Hyde, Wing, & Hsueh, 1986; Paffenbarger, Hyde, Wing, & Steenmetz, 1984; Paffenbarger, Wing, & Hyde, 1978).

Corbin and Pangrazi (1996) adapted Haskell's LPAM for children. The Children's Lifetime Physical Activity Model (CLPAM) encourages children to

accumulate 30 to 60 minutes of MVPA over the course of a day (Corbin & Pangrazi, 1996). Corbin and Pangrazi (1996) suggest that children's MVPA can be accrued, both within and out of the school context.

The impact of physical activity on health appears to have several benefits, many of which do not depend on physical fitness. Physical fitness is influenced by genetic endowment, environmental factors, and general lifestyle behaviors (Bouchard et al., 1994; Pate, Dowda, & Ross, 1991). Evidence points to the need for high intensity training when trying to improve one's health-related physical fitness (Freedson & Rowland, 1992). Physical activity, however, is not associated with heredity, can benefit all youth, positively impacts the health of youth at a lower activity intensity, and is a behavior that can be tracked from childhood to adulthood (Pangrazi & Corbin, 1998).

Physical Activity Measurement

Given the benefits of moderate amounts of physical activity has on one's health, health professionals have been prompted to understand physical activity adherence and its impact on public health. Dishman, Washburn, & Schoeller (2001) have called upon researchers to find methods of physical activity assessment that are unobtrusive, practical to administer, reliable, accurate, and are specific about the type frequency, duration, and intensity of physical activity. Although these researchers concede that no single "gold standard" exists for judging the validity of physical

activity measurement in free-living populations, they outline several popular and appropriate methods to measure physical activity. Dishman, Washburn, & Schoeller (2001) also state two main purposes of physical activity measurement. First, the authors believe that researchers must continue to understand the biological reasons why physical activity protects humans against certain chronic diseases. Second, public health professionals must understand people's motivation to be physically active. The authors summarize several methods of physical activity measurement. These include systematic observation, self-reports, heart rate monitors, pedometers, and accelerometers. Each of these methods will be reviewed in turn, including its advantages and disadvantages for measuring physical activity in free-living populations.

Systematic, or direct, observation is a popular and inexpensive means to collect physical activity behavior. Through systematic observation, humans or machines (e.g., video camcorder) must observe and record a targeted behavior (e.g. MVPA). Its purpose, as noted by Johnston and Pennypacker (1980) "is to arrange conditions so that man or machine will react sensitively to the defined dimensions of the subjects' behavior" (p.146). Although systematic observation techniques can provide specific information about the type of activity, frequency, and duration, it cannot directly measure activity intensity in terms of energy expenditure. A distinct advantage of systematic observation over other methods of physical activity measurement is that it can, through the use of a video camera, provide a permanent visual record of human

behavior. Through this visual record, the rate of energy expenditure can be estimated with specifically designed instruments (e.g., System for Observing Fitness Time Instruction [SOFIT]) and health professionals can determine if the participants being viewed are working at the required amount of energy to receive a health benefit.

Human or machine surveillance of physical activity through systematic observation is practical in most field settings. However, as Dishman, Washburn, and Schoeller (2001) point out, systematic observation is impractical for large-scale epidemiological studies. Furthermore, participants' reactivity to the cameras and/or observers can alter their typical behavior if they are aware they are being observed. However, most experts agree that human behavior will become stable after only a few sessions (van der Mars, 1989a).

Self-report methods to assess the physical activity behaviors of large groups are very popular in the epidemiological and sports psychology literature. Since 1960, over 30 physical activity survey questionnaires have been developed (Dishman, Washburn, & Schoeller, 2001). Although self-report physical activity questionnaires differ greatly on accuracy, they are widely used due to their acceptability by participants, cost, practicality, low interference with participants' daily habits, and ability to provide specific, excluding intensity, activity information (Dishman, Washburn, & Schoeller, 2001).

The challenge of these types of physical activity measurements are their variability in accuracy due to the length of time participants are asked to recall their

physical activity behavior. As the time participants are asked to recall their physical activity engagement lengthens, the greater the potential for gaining a person's true physical activity behavior. However, the greater the time participants are asked to recall also equates to a stronger chance that their memory will be inaccurate (Dishman, Washburn, & Schoeller, 2001). Dishman, Washburn, and Schoeller (2001) point out that "despite their practical appeal, self-ratings of physical activity remain inherently flawed because they depend upon a person's ability and motivation to report accurately.

Heart rate monitors are one of the most common instruments utilized to collect physical activity behavior. Those that measure electrical activity of the heart with chest electrodes are considered the most accurate. Monitors that utilize a photocell to measure opaqueness of blood flow are less valid and considered worthless by some (Dishman, Washburn, & Schoeller, 2001).

Although correlations between heart rate monitors and electrocardiographic recording are quite high (0.95 - 0.97, SEE = 4.7 - 6.3 beats/min^{-1}), converting these data to energy expended is more difficult (Dishman, Washburn, & Schoeller, 2001). To convert heart rate data to energy expenditure values one must develop a linear regression equation between heart rate and rate of oxygen uptake (VO_2). These values tend to be fairly accurate for individual exercise tests (e.g., treadmill running); however, can vary greatly for one individual across different testing days. Therefore,

experts suggest collecting an individual's heart rate data over several testing days utilizing the same exercise test.

Dishman, Washburn, and Schoeller (2001) believe that heart rate monitors can provide accurate and objective data. Recently, Strath et al. (2000) correlated heart rates with VO_2 scores during 15 minutes of physical activity. Results of the study indicated that heart rate data could accurately quantify physical activity behavior. Heart rate monitoring data can be collected from people of all ages with little interruption of their daily living activities. Additionally, compared to other objective instruments (e.g., accelerometers), heart rate monitors are inexpensive and can provide information about the intensity of the activity.

Pedometers are classified under an umbrella type of physical activity measurement device termed a motion sensor or movement counter. These devices have been found to relatively accurate with adult populations when compared to criterion measures of walking or running (Welk et al., 2000). However, Welk et al. (2000) found a very low correlation between average daily step count and self-report methods. Some researchers have discussed concern about the inter-unit reliability of pedometers and suggest more research is needed before these devices can be utilized to collect valid and reliable physical activity behavior (Dishman, Washburn, and Schoeller 2001). Although the confidence in these devices is lacking from a measurement perspective, many experts believe that they are useful for validating self-

report estimates of walking and can be utilized as a motivational device to increase walking behavior (Dishman, Washburn, and Schoeller (2001).

Another device that is used to collect physical activity behavior through changes in motion is called an accelerometer. These are small devices, usually worn on the wrist, hip, or ankle that measure limb acceleration and can provide information about frequency, duration, and intensity of movement. This device provides data in the form of activity counts per given time allotment. The activity count provides researchers with the summed magnitude of the accelerations. A major advantage of these devices over heart rate monitors is that they can collect and store data sequentially over time (Dishman, Washburn, and Schoeller (2001).

The validity, reliability, and calibration of accelerometers have been investigated with promising findings (Nichols et al., 1999). Accelerometers can distinguish between low, moderate, and high levels of physical activity. Energy expenditure estimates can be calculated from these devices. Nichols et al. (1999) found that 90% of the variance in energy expenditure was explained by the accelerometer activity counts.

A major drawback of these devices is cost (approximately $300-500/unit). Other disadvantages of these devices include the inability to collect motion during aquatic activities or static activities that do not require movement around an individual's center of gravity (e.g., cycling or rowing). Still, as the cost of these

devices decreases and their ability to collect multiplane movement data improves, the feasibility of monitoring large-scale populations over several days will increase.

Humans are multidimensional in nature and physical activity is a complex human behavior. No single physical activity measurement tool can capture its complexities. Dishman, Washburn, and Schoeller (2001) suggest that the measurement tool chosen should be based on the research question being asked and the constraints of sample size, timing, setting and budget. These authors suggest that, in research settings, multiple assessment methods should be utilized in order to gain a more global estimate of physical activity.

Systematic Observation and Momentary Time Sampling

Systematic observation "allows a trained person following stated guidelines and procedures to observe, record, and analyze interactions with the assurance that others viewing the same sequence of events would agree with his [or her] recorded data" (Darst, Mancini, & Zakrajsek, 1983, p. 6). According to van der Mars (1989a) systematic observation includes six critical steps: 1) deciding what to observe, 2) developing definitions of the behaviors being observed, 3) selecting the most appropriate observation tactic and determining if there is an existing observation system that fits the need of the observer, 4) establishing observer reliability, 5) making the observation, 6) summarizing and interpreting the collected data.

As mentioned earlier in this review, experts consider systematic observation to be an accurate tool to assess physical activity behavior (Keating, Kulinna, & Silverman, 1999; McKenzie, Sallis, & Nader, 1991; McKenzie et al., 1991; McKenzie, Marshall, Sallis, & Conway, 2000; O'Hara, Baranowski, Simons-Morton; Wilson, & Parcel, 1989; Pope, Coleman, Gonzalez, Barron, & Heath, 2002; Rowe, Schuldheisz, & van der Mars, 1997). Although it has only recently gained attention for this purpose, its beginnings can be traced back to such areas as anthropology, social psychology, clinical psychology, cross-cultural psychology, and classroom teaching (van der Mars, 1989a). In terms of the development of physical activity assessment tools, systematic observation played its most important role in teasing out certain teacher behaviors (e.g., active supervision), as well as student behaviors (e.g., time on task), associated with student achievement.

Because an observation tool has specific coding rules and procedures not ensure its reliability. Several considerations must be made in terms of observer error when collecting data through systematic observation techniques to ensure observer reliability of coded events. The possibility of observer drift, observer expectancies, and observer reactivity must be accounted for. When the observer begins to change coding rules and interpret behavior definitions differently, (s) he is falling victim to observer drift. Causes may include long periods between instrument use, combining or mixing definitions from other instruments, and unclear behavior definitions (van der Mars, 1989b). Observer expectancies, or bias, may occur due to the observer

being aware of investigation's purpose. Finally, knowing that someone else will be checking the observer's reliability may produce observer reactivity. Observer reactivity can be reduced by telling the observer all observations will be screened, performing random interobserver checks, and/or video-taping the sessions so that the observer and his or her assessor will never come in contact.

Even when the above considerations can be accounted for, systematic observation does have several limitations. First, it can only detect observable events, or those events that can be seen or heard. This does mean that such constructs as attitude or emotion cannot be detected through a systematic observation tool (van der Mars, 1989a). Indeed, if these constructs can be categorized and defined, one can assume that these constructs can be witnessed through observable events. For example, if an athlete is consistently rude to his coached during practice, this might be an indication of a poor attitude toward practice.

The second major limitation of systematic observation is the fact that users need to be conscious of the fact that these tools can only provide descriptive data. The user is cautioned when trying to make judgmental or evaluative interpretations as in the example above. As van der Mars (1989a) points out, "the data themselves do not make judgments about how well or poorly the [participant] performed, rather they provide the information for judgments [based on multiple observations] to be made later" (p. 9).

Finally, one must remember that data collected through systematic observation techniques are always contextual. The user must consider the environment, or situation, in which the organism is being observed and to be careful when making generalizations. In the example above, the frequency of the athlete's rude comments can be recorded and used as a piece of evidence supporting a coach's assumption that this athlete has a poor attitude toward practice. However, one must be cautious when interpreting frequency data as it only tells part of the story and does not consider the contextual variables occurring in the athlete's life. It may be that this athlete has had to take on an after school job that conflicts with practice to assist his/her family.

Built into the process of systematic observation are certain recording tactics fundamental to its implementation. These include event recording, interval recording, duration recording, and momentary time sampling. It is the last two tactics that have particular relevance to this investigation, however, a short description is provided for each.

Event recording is the appropriate tactic when investigating behaviors that occur over and over again. Data from this tactic will provide the user with the frequency at which the behavior occurred during a discrete event (i.e., an event with a definite beginning and end) (van der Mars, 1989b). Whether the event is discrete in nature is the critical aspect one must consider when determining if event recording is the appropriate tactic to utilize. The observer must be able to identify a definite

beginning and end. Physical activity examples of events that could be captured using event recording include the number of times participants walked 20 minutes or more during a week or the frequency at which participants rode their bicycle to work during one month.

Although interval recording is sometimes confused with event recording, these are entirely different tactics with different goals and outcomes. Unlike event recording, *interval recording* allows the observer to measure the occurrence or nonoccurrence of behavior within specific periods of time, or intervals. The length of the interval depends on two factors: 1) the experience of the observer and 2) the complexity of the observation system (van der Mars, 1989a). Data from interval recording are reported in percentage of intervals. Calculating the percentage of intervals, based on a 30 second interval, that individuals use the stairs versus the elevator in a busy bank lobby is an example of the use of an interval recording tactic. One unique feature of interval recording is that it provides an estimate of both the frequency and duration of behaviors (van der Mars, 1989b). This to is its biggest disadvantage when compared to event and duration recording as it only provides an estimate of the behavior.

When targeted behaviors last for extended periods of time the appropriate observation tactic to utilize is termed continuous duration recording. *Duration recording* (DR) offers a complete record of behavior and serves as a measure to

compare other estimates of behavior (van der Mars, 1989b). DR represents the "true" state of nature. This tactic might be used to investigate the amount of time students engage in moderate physical activity behavior on a school playground or the amount of time adults spend in occupational activities that might benefit their physical health. Although DR provides a complete record of behavior, this tactic is not practical for behaviors in certain settings. For example, it would not be feasible to track a person's physical activity over the course of one week, or even over the course of one day, through the use of DR. Instead, investigators rely on assessments that sample for targeted behaviors throughout an allotted time. Heart rate monitors, accelerometers, and systematic observation tools can be utilized to sample physical activity behaviors throughout a given time period (e.g., one week).

Similar to interval recording *momentary time sampling* (MTS) provides an estimate of true behavior. MTS allows the user to take samples of behavior during an event (e.g., teaching episode) at the end of each observation interval (e.g., 10 seconds; van der Mars, 1989b). The measured targeted behavior is scored as having occurred if it was present at the end of the observation interval (see Figure 1). Unlike interval recording, MTS procedures can be utilized by teachers during an instructional episode. Test and Heward (1984) outlined several practical advantages of MTS for teachers:

First, it does not require continuous observation of the subject(s); the observer need only "look up" when the interval ends. This means that the teacher can engage in other activities and simultaneously collect data on behaviors of interest. Second, since a teacher can continue his or her regular duties and also collect data, the need for an outside observer may be eliminated. And third, since momentary time sampling does

not require suspension of all other activities in order to measure behavior, data collection often can be conducted over a greater number of days or weeks. (p. 178).

A second major advantage of MTS is that it allows for an outside observer to use the non-observational time to monitor other targeted behaviors (van der Mars, 1989b).

The utility of MTS procedures has been established while sampling both animal and human behavior (Harrop & Daniels, 1986; Leger, 1977; Powell, Martindale, & Kulp, 1975; Powell et al., 1977; Repp et al., 1976; Simpson & Simpson, 1977; Saudargas & Zanolli 1990). Powell, Martindale, and Kulp (1975) targeted the in-seat behavior of a secretary to assist in determining the accuracy of time sampling recordings compared to DR.

In-seat behavior was measured for two 20-minute sessions each morning and afternoon by a video camera. Observations utilizing MTS occurred every 10, 20, 40, 80, 120, 240, 400, and 600 seconds.

Data were analyzed visually to determine the differences between each measure. Results indicated that when MTS procedures were made 10, 20, 40, 80, and 120 seconds there was little difference between the time samples and duration measure. The largest difference observed between these five values and the duration measure was 8 percent. Beyond the 120 second interval larger discrepancies (15% - 74%) were observed.

The authors concluded that the shorter the interval was between observations the greater the correspondence between the two measures. The authors also noted that although the MTS procedure tended to under- and overestimate the true behavior, a

summation of these differences over sessions can "give a result that quite closely approaches a continuous measure" (Powell, Martindale, & Kulp, 1975; p. 467). This point is critical when trying to measure physical activity behavior of children using MTS procedures. Due to children's physical activity patterns during an entire physical education class, their MVPA behavior may be quite high or low depending on such variables as the lesson content, location, and teacher's behavior. Therefore, it is critical for teachers to evaluate their students' MVPA behavior throughout the entire class.

In a similar study, Repp et al. (1976) investigated the accuracy of frequency, interval, and time sampling procedures while observing pseudo-behavior created by electromechanical equipment (i.e., pen deflections on paper). Data were generated by two conditions to simulate true behavior: 1) high, medium, and low rates of responding, and 2) as a constant rate of responding. Two time sampling conditions were investigated. First, all pen deflections produced by the electromechanical equipment for the first 9 minutes and 55 seconds were ignored. Only the last 5 seconds of the 10-minute interval were utilized under this condition. Second, all pen deflections during the first 9 minutes and 50 seconds were ignored and only the last 10 seconds were utilized for the time sampling procedure. These conditions produced 36 and 18 time samples during a 180-minute session, respectively.

When the results from the time sampling condition were compared against an interval recording condition the authors concluded that the time sampling condition

did not produce an accurate representative sample of behavior. This was particularly apparent when the behavior was irregular across the entire session and when it was of medium or high rate. The authors concluded that time sampling of moderate interobservation periods does not properly represent true environmental events. Most importantly, the authors found that there was an interaction between the rate and pattern of responding and the method of data collection.

Although one cannot feel confident about the results of this study due to the lack of a criterion measure of behavior (i.e., duration recording), the authors remind readers of the relationship between rates of responding and the measurement of behavior. This has particular relevance to the study of human physical activity behavior because of the variability in rates of responding depending upon the age demographics under study. For example, we know that adults tend to be moderately active for longer periods of time than children. Children, unlike adults, tend to participate in vigorous activity for short periods of time with frequent periods of rest.

Some evidence demonstrates that MTS procedures provide the most accurate estimation of true behavior when intervals are 60 seconds or less (Powell et al., 1977). Additionally, Leger (1977) found for intervals as large as 120 seconds, MTS produced similar estimates of continuous duration recording when investigating chimpanzee behavior in their natural environment.

Although not directly examining the most accurate length to estimate behavior through MTS, Harrop and Daniels (1986) investigated the accuracy of MTS and

partial interval recording in estimating both absolute behavioral levels and relative change. The authors used a computer to randomly generate runs of pseudo behavior varying in duration and rate. Rates of behavior were classified as low to medium and medium to high. The MTS procedure varied from 20, 40, 60, 80, 100, 120 and 120, 240, 360, 480, 600, 720, respectively. Unlike Repp et al's. (1976) study, Harrop and Daniels (1986) compared MTS data against measures of behavior from DR and interval recording.

Results indicated that in both sampling methods, sensitivity increases with lower rates of responding and longer emitted duration of behavior. Linearity graphs of both sampling methods demonstrated that MTS procedures are far superior to estimates of behavior using the partial interval method, especially for medium to high rates of responding. Graphs indicated that the partial interval method created more systematic error than the MTS procedure and that it tended to underestimate true behavior. The authors concluded that both estimates of behavior were inaccurate, MTS procedures provided the most precise estimation of absolute duration. Additionally, the authors found that MTS is the less sensitive method in terms of recognizing relative change in behavior. However, it did not suffer from as large amounts of systematic error as the partial interval method during the course of the full observation.

Most recently, Saudargas & Zanolli (1990) investigated the accuracy of MTS against DR collected through handheld computers in an elementary school. Twenty-two observations were conducted in elementary schools by one observer who used 15-

second MTS procedure and a second who used a handheld computer to collect DR data. Six student behaviors were targeted and defined using the State Event Classroom Observation System (SECOS), a multiobservation tool used in classroom research to assess children with learning disabilities. The six targeted student behaviors included schoolwork, looking, other activity, child interaction, teacher interaction, and out of seat.

Results demonstrated a close (less than 9%) correspondence between the MTS observation percentage and the DR percentage for each of the six behaviors. The authors found that the 15-second MTS procedure tended not to be as sensitive for low-frequency, short-duration behaviors. This finding has been consistent in other investigations (Harrop & Daniels 1986; Powell, Martindale, & Kulp, 1975; Repp et al. 1976). The authors concluded that MTS estimates can accurately capture behaviors for a wide range of frequencies and durations, and suggested that observers using MTS in a natural setting are able to obtain accurate data.

From these studies it is obvious that one must be cautious when interpreting MTS data, as it only provides an estimate of true behavior (Test & Heward, 1984). When MTS procedures are compared to measures of the same behavior produced by DR, it is possible that they can over- or underestimate the true behavior (Powell, Martindale, & Kulp, 1975, Harrop & Daniels, 1986). In general, underestimation may occur when the intervals are long and the typical duration of the target behavior is short. Overestimation may result when intervals are short and the typical duration of the target behavior is typically long.

Thus, accuracy of MTS procedures depends upon the proper selection of interval length utilized during an observation relative to the typical duration of the target behavior. When the frequency and length of intervals closely match, MTS procedures offer excellent estimates of actual behavior when compared to duration recording (Leger, 1977; Powell, Martindale, & Kulp, 1975; Powell, Martindale, Kulp, Martindale, Bauman, 1977; Simpson & Simpson, 1977). Generally, findings from these investigations have demonstrated that a) MTS provides an accurate estimation of true behavior collected through duration recording in lab settings and b) 120 seconds seems to be the maximal interval length for MTS.

Educational Reform and Authentic Assessment

Educational reform is tough issue facing today's teachers, as they have been asked to find "authentic" ways to assess student learning on an on-going basis (Greenwood & Maheady, 1997, NASPE, 1995; Wood, 1996; 2003). During the past decade, countless physical education research articles, journal monographs, and texts have been dedicated to authentic assessment issues and are evidence of the paradigm shift advocated by teacher educators (Brands, 1996; Cleland & Stevenson, 1997; Melograno, 2000; Richard, Godbout, & Griffin, 2002; Siedentop, 1994; Smith & Cestaro, 1998; Weinberg, 1996; Wood, 1996; 2003). Wood (1996) asserts "physical education may be on the verge of an assessment revolution characterized by greater integration of assessment into the instructional process and development of alternative

assessment strategies" (p. 213). *Alternative assessment* is an umbrella term for those strategies that broadly assess what students know when they exit a grade level and tend to integrate higher order psychomotor, cognitive, and affective objectives under realistic conditions (Wood, 1996). Specific authentic assessment strategies include portfolios, discussions, debates, case studies, student activity logs, and role-playing.

Authentic assessments are a type of alternative assessments and have been described as performance assessments conducted in real-world contexts (Siedentop & Tannehill, 2000; Veal, 1992). Veal (1992) characterized authentic assessment as "regular and ongoing [i.e., formative], [there is] a connection between daily instructional tasks and assessment, the teacher can see the skill that is being evaluated, and there is a connection of skills to real-life situations a learning indicators" (p. 90). Siedentop and Tannehill (2000) concur and define authentic assessment as those events that reflect real life, are performed in realistic settings, and mirror what students do outside of school.

Authentic assessment does pose some practical issues related to reliability, objectivity, and validity. Due to the subjective nature of these assessments, they require a detailed description of scoring criteria and a sufficient amount of practice implementing the processes to maximize the reliability of scoring (Wood, 1996). Authentic assessments must also be shown to have content-related evidence of validity (i.e., measuring important learning outcomes and covering material presented during class) and concurrent-related evidence of validity (i.e., comparing them to tests with established validity; Wood, 1996).

Resulting benefits of this assessment model are numerous as it "tends to engage students in the learning process; allows for the assessment of higher order cognitive processes; permits assessment of a combination of cognitive, psychomotor, and affective behaviors; permits the assessment of what students know and can do in real-life contexts; and provides an assessment method for integrated learning across subject areas" (Wood, 1996, p. 214). It is believed that, through these processes, the authentic assessment model will provide a powerful means to link teaching, learning, and assessment (Wood, 1996).

Greenwood and Maheady (1997) take the usefulness of authentic assessments one important step further for teacher educators. They suggested that the assessment of on-going student behavior be the "gold standard" by which teachers, and teachers educators, make judgments regarding effective teaching. Traditional pre- and posttest types of assessments lack in-depth knowledge of student performance and provide no direct evidence that students have progressed as a function of a teacher's instruction. Greenwood and Maheady (1997) believe that the "inability to document meaningful changes in student performance has…impeded our ability as teachers and teacher educators to identify those instructional arrangements and practices that may be responsible for subsequent changes in learner performance" (p. 266).

Greenwood and Maheady (1997) also contend that the field of teacher education has "failed to adopt an improvement-oriented inquiry approach to teaching" (p. 266). This concept is important as teacher educators ask inservice and preservice teachers to continually reflect on their teaching and consider how it impacts student

growth throughout a unit, not just at the beginning and end. If teacher educators can assist students of teaching be more reflective in this manner, we may empower them to make valuable decisions during teaching and future instruction.

Summary

The benefits of physical activity and its association with reducing all cause morbidity and mortality are well documented. Experts believe that physical education in schools can play an important role in promoting a physically active lifestyle and that elementary schools ought to be the primary location to promote these behaviors. The USHHS department has categorized physical inactivity as a major risk for several chronic diseases and has challenged physical educators to dedicate at least 50% of their class time to moderate to vigorous physical activity (MVPA). Although physical educators might have to make small curricular and pedagogical changes to meet this challenge, they at present have no method of assessing their students' MVPA behavior in a formative and practical manner. Results from investigations demonstrate that MTS procedure, a method utilized in applied behavior analysis to sample behavior, is an accurate method to assess behavior. If teachers are to assess MVPA behavior in authentic formative ways, they must be provided with valid and reliable tools to do so.

Appendix B

OREGON STATE UNIVERSITY
INSTITUTIONAL REVIEW BOARD
312 Kerr Administration Building · Corvallis, Oregon · 97331-2140
E-MAIL: IRB@oregonstate.edu · PHONE: (541) 737-3437 · FAX: (541) 737-3093

REPORT OF REVIEW

TO: Hans van der Mars,
Exercise and Sport Science

RE: Validation of the SOFIT Direct Observation Instrument's Physical Activity Categories Across Varying
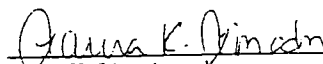Observation Tactics

Protocol No. 2035

The referenced project was reviewed under the guidelines of Oregon State University's Institutional Review
Board (IRB). The IRB has approved the application. This approval will expire on 10/14/03. This new request
was reviewed at the Exempt from Full Board level. A copy of this information will be provided to the full IRB
committee.

- Any proposed change to the approved protocol, informed consent form(s), or testing instrument(s) must be
  submitted using the MODIFICATION REQUEST FORM. Allow sufficient time for review and approval by
  the committee before any changes are implemented. Immediate action may be taken where necessary to
  eliminate apparent hazards to subjects, but this modification to the approved project must be reported
  immediately to the IRB.
- In the event that a human participant in this study experiences an outcome that is not expected and routine
  and that results in bodily injury and/or psychological, emotional, or physical harm or stress, it must be
  reported to the IRB Coordinator within three days of the occurrence using the ADVERSE EVENT FORM.
- If a complaint from a participant is received, you will be contacted for further information.
- Please go to the IRB web site at: http://osu.orst.edu/research/RegulatoryCompliance/HumanSubjects.html
  to access the MODIFICATION REQUEST FORM and the ADVERSE EVENT FORM as needed.

Before the expiration date noted above, a Status Report will be sent to either close or renew this project. It is
imperative that the Status Report is completed and submitted by the due date indicated or the project must be
suspended to be compliant with federal policies.

If you have any questions, please contact the IRB Coordinator at IRB@oregonstate.edu or by phone at
(541) 737-3437.

Laura K. Lincoln   Date: __10/15/02__
Institutional Review Board Coordinator

pc: 2035 file