

**FOURTH AND FIFTH GRADE CHILDREN'S UNDERSTANDING OF
PHYSICAL ACTIVITY: THE DEVELOPMENT OF
THREE ALTERNATIVE ASSESSMENTS**

Natalie Doering

Dissertation submitted to the faculty of the Virginia
Polytechnic Institute and State University in partial
fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Curriculum and Instruction

George Graham, Chair
Leslie Lambert
Susan Magliaro
Jon Poole
Janet Sawyers

June 30, 1997
Blacksburg, Virginia

Keywords: Physical Education, Alternative Assessment, Physical Activity

Fourth And Fifth Grade Children's Understanding Of Physical Activity: The Development Of Three Alternative Assessment Tasks

Natalie Doering

The purpose of this study was to design, pilot, and evaluate three alternative assessments that measured fourth and fifth grade student's understanding of physical activity guidelines as contained in the Surgeon General's Report (USDHHS, 1996). The alternative assessments were: 1. designing a booklet for a peer named Homer, 2. creating a Video Tape Advertisement for students at another school; and 3. coding the perceived intensity level for one minute intervals at an active aerobic type station (Perceived Intensity Level Assessment Task).

Four research questions guided the evaluation of the assessment tasks. They were: 1. Do students who have been taught the physical activity guidelines score differently on the alternative assessment than those who have not been taught the guidelines? 2. Do content experts agree that scores can be used to describe what students have learned (content validity)? 3. Do students find the assessment task worthwhile, enjoyable, and meaningful? 4. Is the alternative assessment prototype "feasible" for a teacher to administer in a regular physical education setting?

Student score results, student self-reflections, and student interview data were used to evaluate the first and third research questions. Teacher interview data and Content Expert validity score results were used to analyze the second and fourth research questions.

Results from this study indicated that the Homer Booklet Assessment Task discriminated between those students who were taught (experienced group) and those who were not taught (inexperienced group). Furthermore, both the Content Experts and the teacher found this assessment to have high content validity and found it to be feasible to use in a regular physical education setting. Although the students enjoyed the Homer Booklet Task they found it to be the least enjoyable of the three assessment tasks. In contrast, students enjoyed the Video Tape Advertisement Task best. Although this assessment task did discriminate between those who were taught and not taught there were a couple of problems with this assessment. The teacher and Content Experts found this task to have feasibility problems and scores did not discriminate for style. Finally, the Perceived Intensity Level Assessment Task did not discriminate between students who were taught and not taught.

TABLE OF CONTENTS

ABSTRACT.....	ii
TABLE OF CONTENTS.....	iii
LIST OF APPENDICES.....	vi
LIST OF TABLES.....	vii
LIST OF FIGURES.....	viii
CHAPTER 1. Introduction.....	1
Rationale for the Study.....	3
Physical Activity.....	3
Health benefits, risks and needs for physical activity.....	3
Children's physical activity levels.....	4
Surgeon General's Report.....	5
Physical education programs and fitness tests.....	6
Alternative Assessment.....	8
Statement of Purpose.....	10
Research Questions.....	11
Significance of Study.....	11
Limitations of Study.....	13
Delimitations of the Study.....	13
Basic Assumptions.....	14
Definition of Terminology.....	14
Summary.....	16
CHAPTER 2. Review Of Literature.....	18
Physical Activity, Health and Well-Being.....	18
Surgeon General's Report	19
Effects of Physical Activity on Health Risks & Benefits.....	22
Children's Physical Activity Patterns.....	23
Physical Education Programs.....	24
Dissatisfaction with Fitness Tests.....	25
Dissatisfaction with Tests and Testing in Education.....	28
Assessing Physical Activity.....	29
Self-Report Instrument and its Limitations.....	31
Assessing Determinants of Physical Activity.....	33
Support from National Documents.....	36
Constructivist Approach.....	38
Alternative Assessments.....	40
CHAPTER 3. Methodology.....	45
Setting.....	46
Participants.....	49

TABLE OF CONTENTS (continued)

CHAPTER 3. Methodology (continued)	
Procedures for Test Development.....	50
Steps for Test Development.....	50
Task Development Process.....	52
Phases of the Study.....	53
Homer Booklet Assessment Task.....	53
Video Tape Assessment Task.....	56
Perceived Intensity Level Assessment.....	59
Data Collection.....	61
Data Analysis.....	62
Data and Researcher Trustworthiness.....	65
Summary.....	66
CHAPTER 4. Results & Discussion.....	67
Homer Booklet Assessment Task.....	67
Homer Booklet Assessment Task Description.....	67
Homer Score Results.....	68
Homer Validity Question/Content Experts.....	71
Homer Meaningfulness Question.....	75
Homer Feasibility Question.....	84
Conclusion and Interpretation of Results.....	85
CHAPTER 5. Results & Discussion.....	87
Video Tape Advertisement Assessment.....	87
Video Tape Advertisement Assessment Task Description.....	87
Video Tape Advertisement Scores.....	88
Video Tape Validity Question/Content Experts.....	97
Student Meaningfulness (Enjoyment) Question.....	101
Video Tape Feasibility Question.....	105
Conclusion and Interpretation of Results.....	108
CHAPTER 6. Results & Discussion.....	110
Perceived Intensity Level Assessment Task.....	110
Perceived Intensity Level (PIL) Assessment Task Description.....	110
Perceived Intensity Level Task Scores.....	118
Perceived Intensity Level Validity Question/Content Experts.....	120
Meaningfulness Question.....	123
Feasibility Question.....	126
Conclusion & Interpretation of Results.....	127
CHAPTER 7. Summary and Conclusion.....	130
Final Survey.....	130
Summary.....	131

TABLE OF CONTENTS (continued)

CHAPTER 7 (continued)	
Conclusions.....	132
Suggestions for further Research.....	133
Major Contributions of this Study.....	133
Implications of Results.....	134
REFERENCES.....	136

LIST OF APPENDICES

APPENDICES.....	146
A Informed Consent.....	147
B Example of Alternative Assessment Tasks.....	149
C Homer Booklet Task.....	151
D Test Construction.....	152
E Content Expert Agenda.....	156
F Teacher Instruction.....	157
G Video Tape Advertisement Assessment Task (Weighted Version)	162
H Student Designed Station Task (Student Directions1).....	165
I Stations Teacher Design.....	169
J Interview Guide Questions for Students.....	170
K Final Survey.....	171
L Student Completed Homer Booklet Assessment Task Examples.....	172

LIST OF TABLES

Table 1	Fifth Grade Homer Booklet Rubric Scores & Percentages	70
Table 2	Fourth Grade Homer Booklet Rubric Scores & Percentages	71
Table 3	Booklet Assessment Overall Task Enjoyment	78
Table 4	Video Tape Advertisement Assessment: Within Group Mean Rubric Content Scores (Experienced Group)	95
Table 5	Video Tape Advertisement Assessment: Within Group Mean Rubric Content Scores (Inexperienced Group)	95
Table 6	Video Tape Advertisement Assessment Overall Task Enjoyment	102
Table 7	Student Self-Match with Adult Rater	119
Table 8	Fourth Grade Self and Partner-Match with Adult Rater Percent and Number of Matches.....	119

LIST OF FIGURES

Figure 1	Student Checklist for Physical Activity Station Assessment	167
Figure 2	Student Checklist 2 for Physical Activity Station Assessment	168
Figure 3	Homer Booklet Assessment Task (Final version)	69
Figure 4	Booklet Mean Rubric Group Scores	70
Figure 5	Content Validity Rubric (Homer Booklet Assessment Task)	75
Figure 6	Student Self-Reflection and Ideas	77
Figure 7	Video Tape Advertisement Task (Student Directions)	89
Figure 8	Video Tape Advertisement Rubric for students	90
Figure 9	Video Tape Advertisement Rubric for the Teacher and Rater	91
Figure 10	Video Mean Rubric Content Scores	93
Figure 11	Video Mean Rubric Style Scores	94
Figure 12	Content Validity Rubric (Video Tape Advertisement Task)	99
Figure 13	Perceived Intensity Level Assessment Task (Teacher Designed-- Final).....	115
Figure 14	Student Circle Coding Sheet--PIL Assessment Task	116
Figure 15	Teacher Designed Station Tasks	117

CHAPTER 1

Introduction

Historically, physical fitness testing has been the primary benchmark used to measure children's fitness levels in physical education (McKenzie & Sallis, 1996). Furthermore, for the past forty years, data collected from these standardized tests have been used to evaluate the effectiveness of physical education programs.

Now, however, physical fitness testing, although still being widely used, is no longer the only method or manner in which to assess the effectiveness of physical education programs. Specifically, experts are promoting the notion that physical activity can contribute to lifelong health, especially reducing cardiovascular disease (CVD). Recent evidence from *Health and Physical Activity: A Report of the Surgeon General* (U. S. Department of Health and Human Services, 1996--USDHHS) indicates that regular participation in moderate levels of physical activity is more easily maintained and ultimately contributes to many health benefits. The USDHHS report is called the Surgeon General's Report throughout this dissertation.

Living healthy lifestyles and making wise health decisions are important to the health of this nation (U. S. Department of Health and Human Services, 1991). Since childhood is considered a favorable time period to establish healthy habits before behaviors of physical inactivity become established as habits, physical education programs can potentially contribute to the promotion of this nation's health by shifting their focus from the importance of physical fitness to promoting regular participation in physical activity for a lifetime (Corbin & Pangrazi, 1992; McKenzie & Sallis, 1996; Simons-Morton, O'Hara, Simons-Morton, & Parcel, 1987).

Previous research on intervention and factors that determine physical activity behavior, have focused on a behavioral conception of learning. From this

perspective, students are seen as passive responders in the process of learning. Yet, many studies indicate that promoting self-management skills are effective. Self-management skills, or taking an active role in one's health can be used to enhance generalization and maintenance of physical activity through a lifetime (e.g., Kanfer & Goldstein, 1980).

Therefore, this study focuses on children's construction of physical activity understanding as related to their personal lives. From this constructivist perspective, knowledge is not transmitted directly from one knower to another, but is actively constructed by the learner (Bruner, 1979; Dewey, 1915/1962; Driver, Asoko, Leach, Mortimer, & Scott, 1994). In this study, learners are viewed as active constructors of knowledge rather than as passive recipients of knowledge (Brown, 1994).

Recent performance assessments in science, mathematics, and writing also view the learner as an active, rather than passive, constructor of knowledge (Shavelson & Baxter, 1992; O'Neil, 1992; Shepard, 1995; Lambert, 1996). From this perspective, in fact, the lines between assessment, teaching, and learning become blurred (Stiggins, 1994; Wiggins, 1993). Prior knowledge becomes a part of the teaching and learning process. Teachers and students report enjoying this type of assessment (Blumberg, Epstein, MacDonald, & Mullis, 1986), and the student becomes the center of focus, as compared with the teacher being the center as the knowledge giver (Sizer, 1991).

Although there are ways to measure physical activity behavior through self-reports, objective measures, and direct observation instruments (Sallis, Buono, Roby, Micale, & Nelson, 1993), we do not have a prototype, or ready-to-use assessment in physical education that measures understanding of how to be physically active for a lifetime from a constructivist perspective. Thus, in this

study, three alternative assessments were developed that measure children's understanding of physical activity. This dissertation describes that process.

This chapter contains an introduction, the rationale for the study; statement of purpose; research questions; significance of the study; limitations, delimitations, and basic assumptions of the study; definition of terminology, and a brief summary of the chapter.

Rationale for the Study

Physical Activity

Health benefits, risks and needs for physical activity.

Scientific evidence from the 1996 Surgeon General's Report has recently summarized the abundant health benefits associated with participation in regular physical activity. Physical activity:

- reduces the risk of premature death in general.
- specifically reduces the risk for cardiovascular disease and colon cancer.
- helps prevent or control high blood pressure, obesity, and diabetes which are also risk factors for cardiovascular disease.
- helps maintain optimal weight, and appears to positively affect body fat distribution.
- appears to build greater bone mass during childhood and early adolescence, and it helps maintain bone mass in adulthood.
- helps prevent or control osteoporosis and prevent back pain.
- increases energy level, improves appearance and self-concept, and increases feeling of well-being (USDHHS, 1996).

In contrast, physical inactivity is a major risk factor for cardiovascular heart disease, and a risk factor for other previously mentioned diseases. Despite all these health benefits, only about half of all American adults participate it moderate or

vigorous physical activities on a regular basis, and about 24% of the population is sedentary (USDHHS, 1996). From these numbers, it is clear that physical inactivity is a serious, nationwide problem, and one that needs attention at a young age (USDHHS, 1991).

Furthermore, there is some evidence that cardiovascular heart disease starts developing early in life. In fact, forty percent of all children ages 5-8 have at least two risk factors for heart disease, and these risk factors generally persist into adulthood (Hales, 1991). Nobody denies that a major benefit of physical activity includes reducing the risk of dying from heart disease, a major cause of death in the United States, though not for children.

Children's physical activity levels.

Recently, there has been an increasing interest in discovering preadolescent children's physical activity patterns. Physical activity is difficult to measure, and there is mixed evidence on children's physical activity levels. Most experts agree, however, that children are the most active group in the population. In fact, compared to adults, preadolescent children have high levels of cardiorespiratory fitness (Krahenbuhl, Skinner, & Kohrt, 1985; Pate & Blair, 1978). Children are also generally active enough to meet health guidelines for physical activity (Sallis, 1993).

Although children seem sufficiently active in youth to meet standards for health, there is a consistent decline in physical activity as children move from childhood to adolescence, especially from 13 - 18 years of age. Males, on average, decrease their levels of physical activity about 2.7 percent per year, while females decrease about 7.4 percent per year (Sallis, 1993).

In fact, increasing physical activity among children is a national health objective (USDHHS, 1991), primarily because of its possible influence on future adult participation in physical activity. One assumption is that active children will

develop skills, enjoyment, and habits that will increase the likelihood that they will be active as adults (Blair, Clark, Cureton, & Powell, 1989; Ross & Gilbert, 1985; Simons-Morton et al., 1987; Simons-Morton, Parcel, O'Hara, Blair, & Pate, 1988).

Therefore, childhood seems to be a favorable time period to establish healthy habits--before behaviors of physical inactivity become established as habits. The assumption is that it is easier to prevent unhealthy habits from developing than to influence them through interventions (USDHHS, 1991). If youngsters can learn how to be physically active before adolescence when activity is known to decline--children may continue their healthy activity levels as they learn how to build physical activity into their daily lives and learn how to become physically active for a lifetime.

Surgeon General's Report.

Participating in physical activity to meet health benefits is different from fitness training for sports, or body building, for example. So, just how much physical activity, and what intensity levels are necessary to obtain health benefits as recommended by the latest research evidence contained in the Surgeon General's Report? This document makes a strong statement about the substantial health benefits that can be obtained from regular participation in moderate physical activity, which is in contrast to previous recommendations requiring vigorous levels of exercise.

First of all, the recent Surgeon General's Report (USDHHS, 1996) recommends that all people in the United States increase their regular physical activity to an appropriate level for them--their capacities, needs, and interests. But ultimately, the report recommends "that all children and adults"... "accumulate at least thirty minutes or more of moderate- intensity physical activity on most, or preferably all, days of the week" (p. 44).

Furthermore, it recognizes the fact that there are important health benefits from intermittent bouts (at least ten minutes at one time) of activity. These ten-minute segments of time can be added together in one day to accumulate that thirty-minute standard. Moderate levels are described as brisk walking, such as a three to four mile per hour walk. What does moderate mean? At this rate, it would take a person twenty minutes to walk one mile going at a pace of three miles per hour. It would take a person fifteen minutes to walk one mile going at a pace of four miles per hour. Therefore, for health purposes--not fitness training--a moderate level of physical activity is the target intensity (Pate et al., 1995; USDHHS, 1996).

From the new guidelines, it is also clear that experts recognize the value of many types (occupational, non occupational, home repair, yard work, walking, swimming, cycling) of physical activities that can be interspersed (ten minutes at a time) throughout the day, instead of having to complete physical activity all at one time.

Physical education programs and fitness tests.

Physical education programs have a unique opportunity to contribute to children's health through providing some daily physical activity, and teaching youngsters how to become physically active for a lifetime. The American College of Sports Medicine (ACSM), the Centers for Disease Control (CDC), and the President's Council on Physical Fitness and Sport (PCPFS), all sponsors of the recent Surgeon General's Report (USDHHS, 1996), recommend educational programs focus on the amount of physical activity necessary to produce health benefits rather than focus on fitness and performance benefits.

One of the main goals of many physical education programs is to promote physical activity for a lifetime. Yet, many children become turned off to participation in physical activity for numerous reasons including: embarrassment,

low skill level, or lack of confidence in their ability (Wankel, 1988). A recent study suggests that many children hate physical fitness tests, especially the mile run. For these children, the mile run is "boring" and "meaningless" (Hopple, 1994).

In the past, physical education programs have tested children on fitness levels using any number of physical fitness tests, including the Presidential Fitness Test, Personal Best, or FITNESSGRAM to name a few. Many teachers do not have a choice as school districts and states require them to test children on fitness levels. Although fitness tests have moved away from skills and more toward health-related components in recent years, leaders in the field still question their validity (Corbin & Pangrazi, 1992; Freedson & Rowland, 1992; Updyke, 1992).

For example, there is little or no evidence of a relationship between pull-ups and health (Corbin & Pangrazi, 1992). Specifically, the ability to lift one's body against gravity is confounded by body weight. Someone might be able to bench press three hundred pounds, yet be unable to lift their body weight--as in doing a pull-up. Does the fact that these people can not do a pull-up mean they do not have sufficient upper arm and shoulder girdle strength, or that they are unhealthy?

Physical education teachers also report dissatisfaction with what fitness tests reflect. Teachers believe that students are learning, yet the use of these tests does not seem to facilitate learning, nor provide good indicators of what was learned (National Association for Sport and Physical Education (NASPE), 1995). It appears that these tests may not match well with what goes on daily in a physical education program. Although physical fitness tests are indeed performance tests, they are also contrived (Meyer, 1992) and not authentic (real-life), and do not offer choices for students (Wiggins, 1993). Physical fitness tests may also be developmentally inappropriate.

Alternative Assessment

Wiggins (1993) reminds us that what we test becomes important. What we test should also be a worthwhile assessment task for students. Focusing on physical activity that might affect the future health of this nation is a worthwhile task. If we begin to assess physical activity, it will become important for learning, and, learning about a healthy lifestyle of physical activity has many benefits to the nation and its economy.

With all the new evidence of the effects of physical activity on health, it makes sense that physical education programs should assess youngsters about physical activity levels that contribute to health--focusing on moderate levels children enjoy and are more likely to maintain. With the shift in focus from fitness to one of health--it makes sense that physical education programs offer alternative assessments that can measure childrens' understanding of physical activity and their ability to apply physical activity to their daily lives. Not only is the subject matter important, but, the level of understanding is also important.

With youngsters, there is evidence that the relationship between physical activity, knowledge, and behavior--specifically knowledge about health effects of physical activity--is not important (O'Connell, Price, Roberts, Jurs, & McKinley, 1985); however, knowledge of how to be physically active may be a significant influence (Gotlieb & Chen, 1985). Knowledge of how to be physically active is a higher level of thinking, because knowledge of "how" is at the utilization level.

Resnick and Resnick (1985), argue that American students are the "most tested but least examined in the world" (p. 5). In fact, most children are tested on average twice per year--thus the term most tested (Neill & Medina, 1989). In addition to classroom standardized test, physical education teachers typically test children twice per year.

To be examined would imply quality--looking for the best. Through standardized tests are we giving students the opportunity to produce their best work? Typically standardized tests are one shot fill-in-the-blank responses. No products or performances are required. Students have no opportunity to correct mistakes or to refine their work. Alternative assessments involve students doing something. If done properly, over time student's have the opportunity to refine and correct their work. Thus, student's are continually constructing and refining their knowledge as they edit or modify their work.

Teaching and practicing higher-order thinking is imperative for youngsters of tomorrow's world (Resnick, 1987; Resnick & Klopfer, 1989; Marzano, 1992). The premise is that information doubles every five years, and there is no way we can passively pour into a person's head all that they need to know. Children must learn how to think, how to learn, how to seek help from others, as well as, work with others.

The importance of higher-order thinking is evident in much of the assessment literature. Resnick (1987) argues that in everyday life--tasks are "messy". Problems are not always clear, and there are many solutions from which to choose. Indeed, some judgments must be made. Therefore, high level, creative, and divergent thinking is necessary. Messy problems may also require a group effort with many people working to solve a task by collaborating on it.

Higher-order thinking through worthy tasks that require synthesizing and analyzing information to show evidence of learning in physical education is important in the messy problems faced in our world. One of which is how to be physically active. Children must decide what activities they would choose to participate in. Selecting activities they enjoy is important as enjoyment has been associated with participation in physical activity for both children and adolescents (Borra, Schwartz, Spain, Natchipolsky, 1995). Therefore, each individual must make

those decisions based on their enjoyment (Stucky-Ropp & DiLorenzo, 1993; Tinsley, Holtgrave, Reise, Erdley, & Cupp, 1995). Also, there are numerous ways to meet the daily thirty minute standard, as well as various intensity levels in which to participate. Thus, multiple decisions must be made to meet the daily physical activity guidelines.

Although, there is no one right way, some ways may be better than others for various reasons such as: availability of equipment and facilities available, time of year, risk of injury, etc. Students have to make these decisions. Giving youngsters opportunities to focus on these types of problems would give them practice with the messy yet worthwhile task of figuring out how to build thirty minutes or more of physical activity into their lives.

Knowing how to be physically active for health purposes is a worthy task with important impact to individuals, families, communities, and the nation as a whole. Activity is real to most children in their daily lives. Choices can be made about a variety of activities to be done in numerous ways to reach the moderate level thirty minute recommendation. Thus, the time has come to develop an assessment that measures this cognitive understanding in young children before they reach adolescence when the levels are known to decline.

Statement of Purpose

The purpose of this study was to design, pilot, and evaluate three alternative assessment prototypes for fourth and fifth grade children that measures their understanding of physical activity guidelines as contained in the 1996 Surgeon General's Report. Understanding was defined as the ability to apply facts, concepts and skills appropriately in new situations (Gardner, 1991).

Research Questions

Four research questions guided this study. They were:

1. Do students who have been taught the physical activity guidelines score differently on the alternative assessment prototype than those who have not been taught the guidelines?
2. Do content experts agree that scores can be used to describe what students have learned (Content Validity).
3. Do students find the assessment tasks worthwhile, enjoyable, and meaningful?
4. Is the alternative assessment prototype "feasible" for a teacher to use in a regular physical education setting?

Significance of Study

Living healthy lifestyles and making wise health decisions (Allensworth, 1996; Hansen, Landsmann, & Monismith, 1996; Hopper, Gruber, Munoz, & MacConnie, 1996) are important to the health of this nation, and physical education programs can play a vital role in this effort. Childhood seems to be a favorable time to learn how to be physically active at levels that confer health benefits. Dissatisfaction with fitness tests and its uses, in conjunction with the national concern for promotion of physically active lifestyles, taken together with early intervention, make this study a need whose time has come.

The design of alternative assessments should focus on process variables and applications of higher-order thinking in children (Hopple, 1995; Mathematical Sciences Education Board (MSEB), 1993; NASPE, 1995; Shavelson & Baxter, 1992; Stiggins, 1994; Wiggins, 1993). Constructing a booklet for someone else to read, and creating a video tape advertisement are two examples that involve higher-order thinking on the student's part.

To focus on process variables, a formative-type evaluation is needed, as evaluation of new interventions and programs should focus on process, so on-going changes can be made (Patton, 1990). Process variables and revisions in assessments, through formative evaluation, are also urged by leaders in alternative assessment development (MSEB, 1993; Stiggins, 1994; Wiggins, 1993; Wiggins, (CLASS), 1995). Therefore, designing and piloting the three alternative assessments for this study involved processes of formative evaluation. However, the implementation phase used summative-like evaluation so judgments could be made.

Furthermore, Herman, Aschbacher, and Winters (1992) argue that alternative assessments should also be developed and piloted in context with students and teachers of the targeted population. Studying the process of the students participating in actual assessments and lessons provides valuable feedback to the development of alternative assessment prototypes. Descriptive research, using a variety of data collection approaches, such as observations and interviews lead to invaluable information about the process of the assessment and children's understanding of physical activity.

The three alternative assessments were developed to measure children's understanding of how to be physically active for a lifetime. The three assessments were designed within the context of one school setting. This study also used a formative evaluation of the process in the development and pilot of the actual assessment prototypes using a variety of data collection approaches which were primarily descriptive in nature. Descriptive statistics were used to describe how groups scored on the alternative assessments and descriptive statistics were used in reporting the Content Expert's and Review Team's overall summative judgment of the instrument and its possible uses for the future.

The detailed methodology of this study is contained in Chapter 3.

Limitations of Study

This study included limitations. They were:

1. The participant's opinions, recommendations for change, and construction of understanding of how to be physically active for a lifetime may not represent other teachers and students in the fourth and fifth grades.
2. The three Content Expert's opinions of the review and critique of the assessment may not represent all physical education experts' opinions.
3. The time and number of pilots used was a limitation in producing a three ready-to-use assessments for fourth and fifth grade youngsters. Even at the end of this study further pilots in various settings will be necessary.

Delimitations of the Study

This study also included delimitations. They were:

1. Participants in the study were delimited to one physical education teacher, two fourth, and two fifth grade classes at one school.
2. The selection of the participants was delimited to those who were enrolled and agreed to participate from each of the four intact classes.
3. The researcher was responsible for the development and initial small group pilot phases. The teacher was responsible for lessons and assessments conducted in the final evaluation phase.
4. Experts were delimited to those pre-selected by the researcher as experts in the field of children's physical education, for their knowledge of physical activity, and knowledge of alternative assessment.

Basic Assumptions

Four basic assumptions guided this study. They were:

1. Student responses reflected the participant's true opinions and recommendations for change.
2. The teacher treated the implementation of the alternative assessment prototype as important in her curriculum.
3. Students viewed the lessons and culminating assessment prototypes as an important part of their school work.
4. The researcher, although intricately involved in all aspects of the study, remained honest and unbiased in reporting the results.

Definition of Terminology

In order to be clear with the concepts and terminology used in this study, definitions and examples are included here. When necessary, please refer back to this section for clarification.

Alternative Assessments are generally performance assessments and involve a student doing something. There are four types of performance assessments. They are: 1) constructed responses, 2) product assessments, 3) performance assessments, and 4) process assessments. These assessment types are further defined in the task development section found in Chapter 3. Typically performance assessments involve the student doing something to apply knowledge such as completing a project (designing a booklet or video or teaching someone else). Alternatives might be booklets, brochures, or portfolios for example (McTighe & Ferrara, 1994; Wood, 1996).

Assessment can be thought of as a comprehensive, multifaceted analysis of performance; it is judgment-based. Assessment involves a variety of techniques. It also relies heavily on observations (of performance), and

human judgment. Furthermore, it involves an integration of (diverse) information in a summary judgments (Cronbach, 1983). Assessment is a process of collecting information about students through informal and formal means. Assessment uses both measurement and evaluation to make judgments about student progress toward a standard (Wood, 1996).

Authentic assessment refers to the context in which the assessment is performed. The more the assessment is like "real-life" the more authentic it is (Meyer, 1992). Authenticity may be real or perceived (Wiggins, 1993).

Evaluation is making judgments regarding the "worth or value of scores" (Wood, 1996, p. 200).

Event tasks are described as short assessments that typically measure a "learnable" piece in a one-day lesson (Lambert, personal communication, October 11, 1996).

Exercise is defined as a subset of physical activity that is planned, structured, and repetitive. It usually has as a final; or an intermediate goal; the improvement, or maintenance of physical fitness or weight control.

Measurement is the process of obtaining a score (Wood, 1996).

Physical activity is defined as any bodily movement produced by skeletal muscles that results in energy expenditure (Caspersen, Powell, & Christenson, 1985). A more reasonable definition of vigorous activity for youngsters would be any activity that is hard enough to make you get tired, breathe harder, or faster than normal, or make you sweat (Sallis, 1993). In the Self-Administered Physical Activity Checklist (SAPAC instrument), physical activity is defined for children as movement of the arms and legs.

Physical fitness can be defined as a set of attributes that are either health--or skill--related. These physical fitness attributes can be measured with specific tests (Caspersen, Powell, & Christenson, 1985).

Prototype is a task, not ready for immediate administration to students. "They are intended to illustrate possible directions for new assessment instruments, not to be an example of a real assessment. Certainly they should be viewed as work in progress, not as fully completed blueprints" (MSEB, 1993, p. 7).

Rubric: A rubric is a scoring guideline for evaluating student work. Rubrics might be qualitative, quantitative or both (Wiggins, 1995).

Test: "A test may be thought of as any standardized procedure for eliciting the kind of behavior we want to observe and measure" (Frederiksen, 1984, p. 199).

Understanding: is the ability to apply (underlined for emphasis) facts, concepts, and skills appropriately in new situations (Gardner, 1991).

Summary

The primary purpose for designing three alternative assessments for physical education programs was to highlight the importance of understanding lifelong participation in physical activity, in contrast to testing physical fitness. The design, pilot, and evaluation of the effectiveness of the alternative assessment was guided by four research questions. The alternative assessment must be feasible to teachers and programs as well as meaningful and worthwhile to students. The other two research questions determined what types of future uses may be possible for the alternative assessment instrument assuming further pilots are conducted in a variety of programs in many geographic areas.

If the instrument is judged as a quality assessment that measures and describes children's understanding of how to be physically active for a lifetime, further implementation of the prototype will be offered for use to various teachers in a variety of programs and possibly with other age groups.

Additionally, many teachers recognize that many children do not enjoy fitness tests--especially the mile run. It was hoped that this assessment would be enjoyable and meaningful for fourth and fifth grade children.

CHAPTER 2

Review Of Literature

Physical activity, physical fitness and health are like a tree. All the parts are related and important, but the essential part of the tree is its roots, for without the roots, the trunk rots, and the leaves wither. Without healthy roots, the tree becomes diseased and dies. In physical education, the root of the tree--the essence, the foundational basic nature, or the sum and substance--is physical activity. Without physical activity, people are at risk for a myriad of diseases. On the other hand, regular participation in physical activity provides many substantial health benefits.

Natalie Doering

The purpose of this study was to design, pilot, and evaluate three alternative assessment prototypes for fourth and fifth grade students measuring their understanding of physical activity as contained in the Surgeon General's Report (USDHHS, 1996).

This chapter details the related research including physical activity, health, and well-being; physical education programs; dissatisfaction with fitness tests; dissatisfaction with standardized tests in general; assessing physical activity; support from national documents; using a constructivist approach; and finally alternative assessments.

Physical Activity, Health and Well-Being.

Physical fitness, physical activity and health are all related to one another, but scientists do not know the full extent of those relationships (McKenzie & Sallis, 1996). True, physical fitness has been part of physical education for many years and is often the component most tested (Hopple, 1995; McKenzie & Sallis, 1996). In fact, when physical education programs are threatened with elimination, many people look for evidence about the importance of physical fitness (McKenzie & Sallis, 1996) because of its relationship to health-related fitness components, including

cardiovascular disease, muscle strength, and endurance (Malina, 1994; Sallis, McKenzie, & Alcaraz, 1993). But, are physical educators barking up the wrong part of the tree?

Surgeon General's Report

The 1996 Surgeon General's Report is the first comprehensive document to address the relationship between physical activity and health. For years, physical educators and health professionals have considered physical activity important in maintaining a healthy lifestyle, and scientists and doctors have known, for years, that substantial benefits can be gained from regular participation in physical activity. Until now, however, no document, has ever reviewed and summarized the existing literature on the role of physical activity in preventing disease.

The 1996 Surgeon General's Report is such a document. The major purpose of this report was to summarize the existing literature on the role of physical activity in preventing disease. This report is the culmination of a collaborative effort by the Office of the Surgeon General who authorized the Centers for Disease Control and Prevention (CDC) and the President's Council on Physical Fitness and Sports (PCPFS) to serve as lead agency for preparing the first report.

Many other agencies collaborated in the final report. Both federal agencies and non-federal partners included:

- Office of the Surgeon General
- Office of Public Health and Science
- Office of Disease Prevention

National Institutes of Health [NIH]

- National Heart, Lung and Blood Institute
- National Institute of Diabetes and Digestive and Kidney Diseases

- National Institute of Arthritis and Musculoskeletal and Skin Diseases.
- American Alliance for Health, Physical Education, Recreation, and Dance
- American College of Sports Medicine,
- American Heart Associations (USDHHS, 1996)

This is an impressive list of federal and non federal agencies collaborating together to share evidence of research as it relates to physical activity and disease control and prevention.

What exactly is physical activity? Physical activity is a broad concept that can be defined as "any bodily movement produced by skeletal muscles that results in energy expenditure" (Caspersen, Powell, & Christenson, 1985, p. 126). Types of physical activity can include daily living or self care, leisure and recreation, and occupation (Ainsworth, Haskell, Leon, et al., 1993). Clearly, physical activities that fall into these broad categories are varied and numerous. Physical activity is not synonymous with structured exercise. Physical activity can be walking, playing volleyball at a picnic, riding one's bicycle to school or work, doing yard or house work, and using stairs versus the elevator at work.

This new understanding of physical activity as it relates to health is a new shift in thinking. Major conclusions of this report confer that people of all ages benefit from regular physical activity. Furthermore, significant health benefits can be obtained by including a moderate amount of physical activity (e.g. 30 minutes of brisk walking or raking leaves, 15 minutes of running, or 45 minutes of playing volleyball) on most, if not all, days of the week (USDHHS, 1996)

Previous (ACSM, 1978) recommendations for developing and maintaining cardiorespiratory fitness and healthy body composition included exercise training of 3-5 days per week, at an intensity level of 60-90 percent of maximal heart rate, a

duration of 15-60 minutes per training session, and rhythmical and aerobic use of large muscle groups through activities such as running or jogging, walking or hiking, swimming, skating, bicycling, rowing, cross country skiing, rope jumping, and various endurance games or sports.

From the descriptions of physical activity and physical fitness, it is clear there has been a shift in thinking and recommendations. Because many important health benefits are linked to thirty minutes of moderate physical activity (USDHHS, 1996), experts agree that for health purposes the regularity of physical activity, maintained over time, most likely produces health benefits, and reduces risks for cardiovascular disease and death (Pate et al., 1995; Sallis, Berry, Broyles, McKenzie, & Nader, 1995; USDHHS, 1996). Simply stated, this means that people no longer have to be vigorously active and exercise three days, at an intense level, in order to be healthy. What a concept!

Recent guidelines are intended to complement new guidelines because any health benefits that are gained depend on a person's current level of activity (Pate et al., 1995; USDHHS, 1996). It stands to reason, therefore, that those people who are not physically active stand to gain the most benefit by beginning to be somewhat physically active while those who already meet the standards can gain some additional health and fitness benefits from becoming more physically active (Pate et al., 1995; USDHHS, 1996).

There is evidence that supports the premise that people who are active for short periods of time (as little as ten minutes at a time--for a combined total of thirty minutes throughout the day) gain significant health benefits just as do those who are active for thirty minutes at one time (Pate, et al., 1995; USDHHS, 1996). Therefore, those people who build physical activity into their daily lives are more likely to be healthy and reduce their risk factors for many diseases.

Effects of Physical Activity on Health Risks & Benefits

Just like a tree can become diseased, so can the human body. Cardiovascular disease (CVD) is the number one killer in America, accounting for half of all adult deaths each year (USDHHS, 1991). Although the first symptom is often a heart attack, the disease starts developing early in life. For example, forty percent of all children between the ages of 5 - 8 have been known to have at least two risk factors for heart disease which generally persist into adulthood (Hales, 1991). Physical inactivity has been identified as one of the primary risk factors and yet approximately one quarter of the adult population is sedentary (USDHHS, 1991).

One of the benefits of physical activity is that, in addition to CVD, it reduces the chance of acquiring other serious diseases. A wide variety of both mental and physical benefits have been linked to moderate levels of physical activity. Specific benefits are listed below.

- People who are even moderately physically active have lower mortality rates than those who are less active.
- Regular physical activity decreases the risk of CVD mortality in general and coronary heart disease mortality in particular.
- Regular physical activity prevents or delays high blood pressure.
- Regular physical activity is associated with a decreased risk of colon cancer.
- Regular physical activity lowers the risk of developing non-insulin-dependent diabetes mellitus.
- Weight-bearing physical activity is necessary for maintaining normal muscle strength, joint structure, and joint function.
- Weight-bearing physical activity is essential for normal skeletal development during childhood and adolescence and for achieving and maintaining peak bone mass in young adults.
- Physical activity may favorably affect body fat distribution.

- Physical activity appears to relieve symptoms of depression and anxiety and improve mood.
- Physical activity appears to improve health-related quality of life by enhancing psychological well-being and by improving physical functioning in persons compromised by poor health.
- Musculoskeletal injuries are believed to be preventable with gradual increases in physical activity.
- It is true that people have died as a result of physical exertion, but, the net effect of regular physical activity is lower risk of death from cardiovascular disease.
[USDHHS., 1996]

It is clear, from this list, that there are serious and substantial benefits to be gained by regular participation in physical activity.

Children's Physical Activity Patterns

Compared to adults, however, children are substantially physically fit (Simons-Morton et al., 1987). In fact, all available data support the conclusion that children are the most active and most fit segment of the U. S. population (Blair, Clark, Cureton, & Powell, 1989; Simons-Morton et al., 1987). Getty, (1995) and Slep and Warburton, (1992) report that children seem to be born with an instinctive love of physical activity. However, this instinctive love is gradually eroded because of unpleasant experiences caused by embarrassment, pain, and failure (Hopple, 1995; Wankel, 1988).

As children move into adolescence, there is a consistent decline in physical activity (USDHHS, 1996). On average, males decrease physical activity 2.7% per year, and females decrease physical activity about 7.4% per year (Sallis, 1993). This decrease is most marked from 13-18 years of age (McKenzie & Sallis, 1996; Sallis, & McKenzie, 1991). From this evidence, it makes sense that intervention programs

for youngsters need to focus on children learning how to enjoy physical activity and stay active for a lifetime, rather than being fit during childhood.

Physical Education Programs

Recently, there has been a shift in thinking on the subject of physical fitness versus physical activity. In fact, many people (Blair, 1992; Corbin & Pangrazi, 1992; Sallis et al., 1992; Sallis & McKenzie, 1991; Simons-Morton, et al., 1987) are suggesting physical education programs become more actively involved in promoting lifetime activity for all people, and Simons-Morton, et al. (1987) specifically recommend that children's programs promote regular physical activity habits that carry over to adulthood, rather than emphasize physical fitness.

Unquestionably, a growing number of health and physical education professionals are increasing their efforts to find methods of getting more people physically active (Corbin & Pangrazi, 1992; Sallis et al., 1992; Simons-Morton et al., 1987). Since children participate in physical education for nine or ten years of their lives, a logical method would be to encourage these programs to promote lifetime physical activity.

Studies, however, are inconclusive on the relationship between childhood physical activity into adulthood (Brill, Burkhalter, Kohl, Blair, & Goodyear, 1989; Powell & Dysinger, 1987). There is some evidence, however, that childhood participation in organized school sports is a precursor of adult physical activity (Brunner, 1969; Sallis & McKenzie, 1991). Not all children participate in organized sports, but most children participate in physical education for many years. Therefore, a shift in thinking to promote lifetime participation in physical activity, in physical education programs, would theoretically provide all children with the knowledge for how to be healthy through physical activity for a lifetime.

The basic assumption is that active children will develop skills, enjoyment, and habits that will increase their likelihood to be active adults (Blair, et al., 1989; Ross & Gilbert, 1985; Simons-Morton, et al., 1987; Simons-Morton, et al., 1988). Physical education programs can help youngsters develop and improve skills. Programs can also promote the notion of enjoyment so that youngsters might continue their pursuits into adulthood.

Because the major impact of physical activity on health is becoming more recognized by health professionals (Gutin, Manos, & Strong, 1992; Strong, 1990; Strong & Wilmore, 1988), and physical education experts (Blair, 1992; Corbin & Pangrazi, 1992; Freedson & Rowland, 1992; Updyke, 1992), a deeper look at assessment in physical activity--in contrast to physical fitness tests--is needed for physical education programs.

Dissatisfaction with Fitness Tests

Unquestionably, there are many concerns with standardized physical fitness tests. Specifically, validity and reliability issues, norm cut-off points and criterion standards, health-relatedness issues, test uses, and children's understanding of the test's purposes are concerns highlighted in this section.

Validity and reliability questions are being raised (Safrit & Looney, 1992; Woods, Pate, & Burgess, 1992) about standardized physical fitness tests. For example, norm-referenced tests scores appear to have arbitrary cut-off points (Updyke, 1992). If that is indeed true, how valid are the scores interpretations?

Criterion-referenced fitness tests have similar problems. One of the major criticisms of criterion-referenced tests is that standards are not well-documented (Blair, 1992; Updyke, 1992). For example, the mile run scores indicate that a higher percentage of both boys and girls fail to meet the standards than the percentage of adults who fail to meet the standards. This raises serious questions about the

standards (Blair, 1992), when there is evidence that children are more fit than adults (Blair, et al., 1989; Simons-Morton et al., 1987).

Furthermore, there are questions concerning the health-relatedness of fitness tests. Because there is no evidence that low strength is related to reduced health (Updyke, 1992), questions are raised when muscle strength is used as an indicator of health. There are also concerns with the meaning of scores (Blair, 1992; Fox & Biddle, 1992; Freedson & Rowland, 1992) and their uses (Fox & Biddle, 1988).

As criticisms about the use of tests (scores rarely informed instruction and learning) were raised about fitness tests, new tests such as the Prudential FITNESSGRAM were developed to include comprehensive programs that include goal-setting. In theory, these improvements should help students acquire knowledge about the meaning of scores, increase motivation, and connect assessment more closely with instruction. When used properly, these tests should have significant potential. However, these tests focus specifically on physical fitness and not physical activity.

Furthermore, any test scores are influenced by affective factors such as self-efficacy and motivation; therefore, reliability becomes a concern when young children take any test (Hopple, 1995; Wiggins, 1993). There is also evidence that as a result of having had to take fitness tests through the school years (Hopple, 1995), youngsters may become turned off to physical activity.

Hopple (1995) interviewed fourth and fifth grade children to gain insights about what they thought, felt, and knew about one physical fitness test battery and the items in that battery. Most children in this study did not have a clear understanding of why they take the mile run test, for example. Furthermore, students disliked taking the mile run test. In fact, many students, if given the chance, would change the mile run to make it more fun .

Hopple (1995) wrote:

...although these insights are representative of only a small number of students and teachers at two schools, it is likely that they are not unique to just those who took part in this study. It therefore becomes germane to examine issues surrounding these results to determine how fitness testing experiences can be made as meaningful and relevant for students as possible. (p. 415)

Conclusions drawn by Hopple (1995) indicate that fitness tests are contrived, and not authentic. Assessments that measure physical activity may not be any more authentic, or less contrived, than fitness tests. However, the premise of connecting instruction more closely with what happens outside of school (authenticity) becomes more possible through physical activity than fitness.

Another problem is that fitness tests are only weakly and partially correlated to physical activity (McKenzie, et al., 1996; Pate, Dowda, & Ross, 1990). Yet, as previously discussed, physical activity produces health benefits. Since it can be argued that physical activity is an essence of physical education--and a strong case has been made for that--and since fitness tests can not measure youngster's understanding of physical activity (nor are they meant to) and how that relates to their daily lives--feasible alternatives that measure physical activity should be developed.

What students might learn about how to stay physically active through their adolescent and adulthood years could become very important for the future public health of American people. Only a long-term study could measure long-term effects, which is beyond the scope of this study. This study, however, can be a beginning in assessing youngster's understanding of physical activity according the Surgeon General's Report. The alternative assessments would simply become another option, an alternative to fitness tests, that practitioners may, or may not,

choose to use to document what students know and are able to do as a result of their programs.

Dissatisfaction with Tests and Testing in Education

Physical education teachers and researchers are not alone in their dissatisfaction with tests. In education in general, many people are questioning testing practices. Lauren and Daniel Resnick (1985) intimate that American students are the "most tested but least examined in the world" (p. 5). In fact, standardized testing has been on the rise in the past ten years. For example, Neill and Medina (1989) conservatively estimate that over 105 million standardized tests were administered in the 1886-87 school year to 39.8 million students. This number indicates that each student takes more than two standardized tests per year. Indeed, most students in physical education alone are tested in fitness twice a year from the fourth grade through the end of their years in physical education programs.

So, how do we know what our students know and are able to do? Where is the evidence of learning? Typically, it is through testing that some of those judgments are made. Yet, "a test by its design, is an artifice whose audience is an outsider, whose purpose is ranking, and whose methods are reductionist and insensitive" (Wiggins, 1993, p. 7). In fact, a test "may be thought of as any standardized procedure for eliciting the kind of behavior we want to observe and measure" (p. 13).

Testing usually occurs after teaching to measure what a student knows after what is taught. Typically testing is seen as separate from learning. In fact, testing to see how students do after teaching is more for other people--not the performer (Wiggins, 1993). When we use tests only to measure, and not help a child learn, the child is treated as an object by any test (Wolf, Bixby, Glen, & Gardner, 1991). Furthermore, tests generally provide relative comparisons of students against each

other by the "normal curve" rather than as a comparison to a standard. In addition, tests generally only measure what is easy to measure, and validity typically gets sacrificed for reliability of results, efficiency of scores, and ranking (Wolf, et al., 1991).

Teaching, learning, and assessment should help children learn and become better at the specific goal, not just measure them compared to a standard or give them a grade. Assessing understanding of physical activity might help students learn how to be physically active for a lifetime.

Assessing Physical Activity

Researchers know more about adult physical activity patterns than they do about young children's patterns. Although not much is known, research has revealed some patterns. For example, there is some evidence that a small percentage of children may not be sufficiently active for health purposes (Simons-Morton et al., 1988). Children may not be active for sufficient numbers of minutes at one time--ten minutes or greater (Sleap & Warburton, 1992). But, many children are active for a sufficient number of minutes each day (Pate, Long, & Heath, 1994; Ross & Gilbert, 1985).

Researchers want to know more about children's physical activity patterns and continue to experiment with various ways to measure physical activity in children. Dozens of methods have been developed for this purpose, but, a "gold standard" has yet to be discovered (LaPorte, Montoye, & Caspersen, 1985; Saris, 1986; Sallis et al., 1996). One of the primary problems of accurately measuring physical activity behavior is the variability experienced in activity from day-to-day. Although a child may be active one day, he/she may not be active in the same habitual manner each day.

Objective measures of physical activity, such as direct observations, heart rate monitoring (Freedson, 1989), motion sensors (Sallis, et al., 1993), and caltrac accelerometer (Sallis, Buono, Roby, Carlson, & Nelson, 1990) are difficult to use with children (Sallis et al., 1993). Furthermore, instruments such as accelerometers and heart rate monitors are expensive and not sensitive to activities such as riding a bicycle and swimming, which many children participate in on a daily basis (Sallis et al., 1993).

Direct observation involves observers observing and recording the behavior in question (LaPorte et al., 1985). Continuous observation is the most discriminating method used in the assessment of children's physical activity levels (Simons-Morton, et al., 1987). However, because of the number of people needed to observe, or rate, each child, this method seems impractical in large scale studies (Freedson & Rowland, 1992). Needing many raters could become cost prohibitive for many school districts.

Observational coding can be done using interval, or time coding of the students. One way to lower costs for training adult coders would be for youngsters to peer rate each other. However, from all the studies reviewed for this research topic, indications are that this method has not been used. Children coding each other could raise reliability issues in the results, but might be considered a feasible option for teachers in individual school settings.

Many self-report measures for fourth and fifth grade students have been found to have at least modest levels of reliability and validity as compared to objective measures (Baranowski, Dworkin, Cieslik, Hooks, Clearman, Ray, et al., 1984; Janz, Witt, & Mahoney, 1995; McKenzie & Sallis, 1996; Myers, Strikmiller, Webber, & Berenson, 1996; McKenzie et al., 1996; Sallis et al., 1993; Simons-Morton, Taylor, & Huang, 1994). The general consensus from these researchers is that self-report measures can be used with moderate degrees of success, but with some

limitations, in large studies. They are fairly easy to administer and are low in cost compared with using the expensive equipment required for obtaining objective measures. Therefore, large-scale epidemiological studies often use self-reported evaluations (Freedson, 1991).

Self-Report Instrument and its Limitations

Again, no self-report can be considered a "gold standard" of true physical activity, and age appears to be strongly related to the quality of physical activity recall data. Furthermore, children younger than age ten cannot be expected to provide useful information on either recalls or diaries. Additionally, measures that reduce the delay between the activity and its report are generally found to be more valid (Sallis, 1993).

One of the problems with using self-reports with young children is that recall of physical activity is a complex cognitive task; (Baranowski, Bouchard, & Bar-Or, 1992) therefore, the time delay between "doing" and reporting becomes critical (Sallis, 1993). The longer the delay, the more difficult it becomes for children to remember.

Children's understanding of time has also been extensively researched. Understanding exact time is quite complex for young children (Friedman, 1982). Therefore, children's understanding of time, in this or any study, could become a confounding variable.

In previous studies using self-reports, discussions of familiar activities helped provide children with a frame of reference for time. The length of television shows, a favorite cartoon, or how long it takes to brush their teeth (McKenzie et al., 1996; Myers, et al., 1996; Sallis et al., 1993; Sallis et al., 1996), helped children understand longer than five minutes, or shorter than five minutes.

Strategies such as breaking up the day into a three-part segment has been successfully used to provide children with a way to understand their day (Baranowski et al., 1984). Previous studies (Mckenzie et al., 1996; Myers, et al., 1996; Sallis et al., 1996) used three segments--before school, during school, and after school-- to help children recall their previous day's activities. The aforementioned studies used the Self-Administered Physical Activity Checklist (SAPAC).

The SAPAC instrument appears to be useful as a measure of relative amounts of physical activity, but not absolute minutes. It is also a useful measure for relative intensity level. There is moderate support for validity in the SAPAC instrument for all gender and ethnic subgroups (Sallis et al., 1996). Furthermore, the SAPAC instrument reports higher validity correlation's than other published self-reports used with children the same age (Sallis et al., 1996).

Further support for use of the SAPAC instrument comes from personal electronic communications to Jim Sallis (September 3, 1996), and Thom McKenzie (September 3, 1996). Both researchers are experts in the field of physical activity and continue to work with youngsters in physical education programs, but because the instrument has not been validated with fourth grade children, this researcher asked about its use with these youngsters. Sallis replied: "It (SAPAC) was validated on 5th grade students, and I would expect [sic] to work less well on 4th graders" (September 3, 1996).

Despite its limitations, ideas taken from the Self-Administered Physical Activity Checklist (SAPAC), instrument were used for this study. Furthermore, interval-type coding by adult raters and children was used in this study. Children were taught about the three part segmented day; before, during, and after school. Children were also asked to think of thirty minutes as compared to their favorite television show. Children were also taught about intensity level as Hard, Easy, or Medium. The notion of being active at moderate levels (most of the time, some of

the time, and none of the time) was found to be difficult with youngsters in this study.

The purpose for designing three alternative assessments is to provide practitioners with choices that are acceptable and feasible ways to document student learning in physical education that fitness tests scores do not document. Reporting accurate amounts of physical activity is not the primary purpose of this assessment; gaining insights into student's understanding of physical activity is.

Time studies in physical education indicate that students are only active about 35% of the time (Faucette, et. al, 1995; McKenzie, et al., 1995; Sallis & McKenzie, 1991). Most students in elementary school have physical education class for thirty minutes at a time, and approximately one hundred minutes per week (Ross & Gilbert, 1985; Ross & Pate, 1987). Therefore, children need to learn how to be physically active, and how to take responsibility for their own healthy lifestyles. If physical education teachers can think of a thirty-minute physical activity session as being one set of a ten-minute interval of physical activity, physical education programs could provide one third of the necessary time. Children could provide the other two thirds before and after school.

Physical education programs typically cannot provide enough activity time during physical education to meet health requirements, let alone reach all other outcome objectives (McKenzie & Sallis, 1996). Therefore, children's activity time outside of school becomes important. It, therefore, seems important to measure children's understanding of "how to be healthy" for a lifetime to meet daily physical activity requirements, including time inside and outside of school.

Assessing Determinants of Physical Activity

Scientists have come to realize that participation in physical activity is influenced by a variety of factors. Determinants, or factors, that influence people to

be physically active or serve as deterrents for people not to be active, become important to this study in the lessons where students will construct their own meanings of physical activity. Research evidence indicates that no one variable, or category of variables, account for most of the variance in children's physical activity. In fact, different variables may be important to different groups (Sallis et al., 1992).

Many believe a major application to public health is that knowledge of determinants can guide interventions. A primary goal of health-related physical education programs is to guide youngsters to be physically active for a lifetime (McKenzie & Sallis, 1996). Providing learning opportunities for children to make personal meanings about physical activity, and providing youngsters with information on how to be physically active in their daily lives, is one way to strive for this goal.

The reason for the focus on how to be physically active comes from studies on the relationships between physical activity, knowledge, and behavior in youth. These studies suggest that knowledge about health effects of physical activity is not important (O'Connell, et al., 1985). However, knowledge of how to be physically active may be a significant influence (Gottlieb & Chen, 1985).

Physical activity research with adults indicates that knowledge of the benefits of physical activity on health does not change behavior. So, the fact that knowledge about health is not important to children is not surprising. This is also consistent with other health-related behaviors. For example, even though people know they should stop smoking, wear their seat belts, exercise, and/or eat healthy foods etc., they don't (Strouse & Fabes, 1985). We, as educators, must realize that no one person can control another person's choices or decisions about physical activity. What we can do, as teachers, is help children understand how to be

physically active, and let them make their own informed decisions with the knowledge they have, or acquire.

In this study, knowledge of how to be physically active includes one major determinant that has been found to be significant with young children. That is the importance of choosing activities children enjoy. The major reason children engage in physical activity is because of the enjoyment (Borra et al., 1995). Indeed, enjoyment has been positively associated with physical activity among both children and adolescents (Stucky-Ropp & DeLorenzo, 1993; Tinsley et al., 1995).

Outside of school, children can control their participation in activities they enjoy that are "fun" for them. In this study, children also had to have some understanding of intensity level and time, which they experienced in two lessons through a variety of activities and the three assessments themselves. It was believed that these experiences would help children better construct their own understanding of how to be physically active for health for a lifetime. In other studies, children have been able to provide relative understanding of intensity (vigorous, moderate, and light), with vigorous being the easiest to accurately recall (Sallis et al., 1996).

Four guidelines from the Surgeon's General Report (USDHHS, 1996) were important in understanding how to be physically active for a lifetime. These criteria were: 1) a person should be physically active most days of the week; 2) at a moderate intensity level; 3) for thirty minutes of accumulated activity time. 4) Children were also taught to select enjoyable activities that are "fun for you". These criteria were taught to the "experienced" group.

Can students accurately inform someone else about how to be physically active for a lifetime? Synthesizing this information requires higher-order thinking, and involves "messy tasks". The ability to apply physical activity understanding was measured through three alternative assessments. They were: 1) Designing a

booklet for Homer (a student of similar age) how to be physically active for a lifetime. 2) Creating a Video Tape Advertisement for children their age at another school. 3) Self coding and self and partner coding of perceived intensity level worked during aerobic type station activities (Perceived Intensity Level Assessment Task). Details of the three assessments are included in Chapters 4, 5, and 6.

Support from National Documents

Further support for the importance of understanding, and being able to apply physical activity to daily living, comes from two national physical education documents: *Outcomes of Quality Physical Education Programs* (NASPE, 1992), and *Moving Into The Future: National Standards for Physical Education* (NASPE, 1995). These documents support the notions of meaningfulness; authenticity; connecting teaching, learning, and assessment; giving children more responsibility for learning; and focusing on essential outcomes. These documents were written by the National Association for Sport and Physical Education (NASPE).

These documents were written from a constructivist perspective. Authors define what a physically educated person should know and be able to do in a developmentally appropriate physical education program. Outcomes are written for every other year and/or exit grade levels (Lambert, 1996). The focus is clearly on the learner, versus the teacher. The National Council of Teachers of Mathematics and the National Science Teachers Association are also arriving at goals for the year 2000, and beyond, that clearly view the learner as one who constructs their own knowledge.

Authors of these physical education documents--experts in the field--clearly recognize physical activity as an important outcome for physical education. The first document outlined twenty outcomes, grouped into five categories, that a

physically educated person should know, and be able to do, as a result of being in a quality physical education program. Three of the categories in the first document speak directly to physical activity, and one indirectly. A physically educated person:

- has learned skills necessary to perform a variety of physical activities,
 - is physically fit,
 - participates regularly in physical activity,
 - knows the implications of, and benefits from, involvement in physical activities, and
 - values physical activity and its contributions to a healthy lifestyle.
- (NASPE, 1992)

The second document was a result of NASPE commissioning a Standards and Assessment Task Force as a direct result of the "Goals 2000: Educate America Act" (Wood, 1996), which was a call from the Federal government for core subject areas to voluntarily establish standards and assessments. Even though physical education was not recognized as one of the eight core subject areas, NASPE responded. Authors of the document narrowed the twenty outcomes from the previous document to seven content standards. Two of the seven standards are directly related to physical activity, and it could be argued that all others are indirectly related to successful participation in physical activity.

1. Demonstrates competency in many movement forms and proficiency in a few movement forms.
2. Applies movement concepts and principles to the learning and development of motor skills.
3. Exhibits a physically active lifestyle.
4. Achieves and maintains a health-enhancing level of physical fitness.

5. Demonstrates responsible personal and social behavior in physical activity settings.
6. Demonstrates understanding and respect for differences among people in physical activity settings.
7. Understands that physical activity provides opportunities for enjoyment, challenge, self-expression, and social interaction. (NASPE, 1995)

Constructivist Approach

The theoretical foundation for this study comes from a constructivist approach to learning. In contrast, most studies of physical activity behavior use a behaviorist approach. It is not this researcher's premise to follow this lead. From this constructivist perspective, "knowledge is not transmitted directly from one knower to another, but is actively built up by the learner" (Driver et al., 1994, p. 5). Support for this type of learning comes from the field of psychology and cognition where learners are viewed as active constructors, rather than as passive recipients of knowledge (Brown, 1994). From this premise, children actively build up knowledge by "doing" or engaging in activity, versus passively absorbing knowledge (Dewey, 1915/1962; Bruner, 1979). In this study, it is hoped that students will construct their own personal meanings about physical activity participation both inside and outside of school.

Further support for the constructivist perspective comes from the classroom. The latest hands-on performance assessments in science, mathematics, and writing are based from this perspective (Lambert, 1996; O'Neil, 1992; Shavelson & Baxter, 1992; Shepard, 1995). From this perspective, teaching, learning, and assessment are linked more closely together, and the lines between assessment and learning become blurred (Stiggins, 1994; Wiggins, 1993). Prior knowledge becomes a part of

the teaching and learning process, and at least one study has found that both teachers and students enjoy this type of assessment (Blumberg, et al., 1986).

In constructivist learning, there is an assumption that children can gain understandings about any subject matter, at some level, if they are given the experiences and opportunities to construct their own knowledge through guided experiences and activity. Within this theory, it is assumed that children can learn any subject content truthfully at any age (Bruner, 1979). Bruner states:

One of the conclusions of the 1959 Woods Hole Conference of the National Academy of Sciences on curriculum in science was that any subject can be taught to anybody at any age in some form that is honest. It is a brave assertion, and the evidence on the whole is all on its side. At least there is no evidence to contradict it. (p.108)

Another tenet of a constructivist approach is to promote higher-order thinking. How do children understand physical activity, and how can they apply that knowledge to their daily lives? In this study, children were asked to synthesize and analyze information, which according to Bloom's taxonomy (1956), involves higher-order thinking. As information about physical activity is evolving, and as researcher's continue to explore this line of research, using higher-order thinking, as it relates to physical activity, is important. Therefore, as researchers continue to expand the knowledge base about effects of physical activity on health, it will be important for children (future adults) to be able to incorporate future knowledge and information with their current understanding.

Another argument for using the constructivist approach is that information doubles every five years. Realistically, there is no way a person can begin to know all there is in the world to know, or even all there is to know about a given topic (Marzano, 1992). Therefore, it becomes important to focus instruction on what is

truly an important outcome. It has been said that if we test it--it becomes important (Wiggins, 1993). If it becomes important--students will focus on it.

As was discussed in the physical activity section, understanding how to be physically active might contribute to healthy living. This is important to public health care and costs. Therefore, it would seem that understanding and applying this knowledge is an important outcome in physical education.

Alternative Assessments

A current reform movement in education questions the validity of tests and, at the same time, encourages schools, administrators, and teachers to move in the direction of using alternative, authentic, and/or performance assessments. Perrone (1991) prefers to use the term *different* assessments, instead of the label, alternative assessments. Whatever terminology we use, what is important is that these assessments can get closer to student learning, and are related to performance and understanding. It makes sense that schooling should focus on student learning, understanding, and performance. For this study, Gardner's (1991) definition of understanding was used. That is: understanding is the ability to apply facts, concepts, and skills, appropriately in new situations (Gardner, 1991).

Leaders in assessment argue that alternative, performance, and authentic assessments can promote and measure higher-order thinking, processes of learning, and deeper, meaningful understandings. These are what Bruner (1979) and Wiggins (1993) call the essences of subject matter knowledge. It was argued previously that physical activity might be considered an essence of physical education. We can not know from fitness tests what students know and are able to do in regards to physical activity. Physical activity contributes to health benefits, reduces risks for numerous diseases and is easier to maintain in daily lives than

fitness. Therefore, it becomes necessary to develop alternatives to measure that knowing.

According to Wiggins (1993), "assessment done properly should begin conversations about performance, not end them" (p. 13.). Fitness results are not often used this way in physical education. Hopefully, with a focus on understanding of physical activity, in and outside of school, and using alternative assessments, teachers and students will begin, and hopefully continue, conversations about physical activity--an essence of physical education.

It is increasingly common in physical education for teachers to use fitness test results for grades, and for administrators to use fitness test results to eliminate programs deemed ineffective as a result of fitness test scores (personal communication with George Graham, 1996). Redirecting our focus to physical activity may not save programs, but might help administrators and policy makers realize the importance of physical education in promoting the public health of this nation.

Alternative assessments that match well with constructivist learning are being developed in many subject areas. According to a 1990 survey by the Center of Research on Evaluation, Standards, and Student Testing at the University of California--Los Angeles (O'Neil, 1992), more than half the States are using, implementing, or exploring the idea of using performance assessments. Writing process and whole language led the way in constructivist perspectives. Many states, including Florida, use writing process in their standardized tests for fourth and eighth grade students.

Other core subject areas are following suit. Hands-on science assessments in California, and at the national level, are being created, piloted and evaluated (Blumberg, et al., 1986; Shavelson & Baxter, 1992; Stecher & Klein, 1996). In mathematics, performance assessments that allow for more than one means to

arrive at a solution are also being created (Blumberg, et al., 1986; MSEB, 1993; Shepard, 1995). As noted before, in real life--problems are messy. There may not be a single, clear, best answer to a particular problem, and real life problems may require synthesis of information. Using divergent thinking, and coming up with divergent responses, may become necessary. In school, in mathematics and science, or in creating a dance, each student may provide a unique response or process to the questions or problems posed.

The alternatives being designed for physical education in this study are certainly messy, as each child will respond differently when creating a booklet for a hypothetical peer named Homer. Furthermore, the Video Tape Advertisement will be different for each group as there are various ways to arrive at the recommended guidelines, including intensity and time, and activities to include.

Tasks or assessments that require divergent processes or responses are not easy to measure. For real student learning to occur, teacher judgment, and even help during the test, may be needed. Therefore, these type assessments rely on human judgments in contrast to the easy-to-measure multiple choice tests, which remove human judgment and are impersonal (Wiggins, 1993). Therefore, rubrics were developed for the Homer Booklet Assessment Task and the Video Tape Advertisement Assessment.

Wiggins (1993) notes that "an assessment is a comprehensive, multifaceted analysis of performance; it must be judgment based and personal" (p. 13). Assessment, done properly, uses a variety of techniques and integrates a diverse amount of information into a summary judgment. Even standardized tests may have their place in the school setting, but no single test should be used to make judgments about a single child (Cronbach, 1983).

To develop an educational system based on the premise that 'all children will learn', we need assessment systems that treat each

student with greater respect, assume greater promise about gains (and seek to measure them), and offer more worthy tasks and helpful feedback than are provided in our culture of one-shot, 'secure' testing. (Wiggins, 1993, p. 6.)

One reason for developing and using alternative assessments is to try and link assessment more closely with instruction (Stiggins, 1994). Teachers in schools assess youngsters informally all the time through observations, checklists, and performances such as oral reading. In physical education, for example, some teachers use checklists to keep track of student's development in various motor skills.

These types of observations are made based on professional judgments compared to a standard of the process of what the skill should look like. Physical education teachers are trained to become effective observers of specific skills. Yet, these observations may not be considered as valid--to districts and administrators--as the "outside" fitness tests. The reform movement in assessment attempts to give power back to the teacher as someone who is able to know and make accurate judgments about what children know and can do in their classroom.

Another reason for developing alternative assessments is to encourage changes in the curriculum. Changing the focus in assessment from fitness to physical activity might encourage physical education teachers to focus instruction on student's understanding of physical activity in their daily lives outside of school. One result could be the better health of our nation.

Can assessments that measure student understandings about physical activity, and the ability to apply physical activity guidelines to their daily lives, be designed as options for practitioners to use in conjunction with, or in addition to, fitness tests? Can more worthy tasks be designed for students? Can performance

and cognitive assessments be combined that would be better indicators of student learning in physical education?

How can physical educators design more appropriate alternative assessments in physical education? Designing three alternative assessments based on the recent Surgeon General's Report--which focuses on physical activity as it relates to health, and not fitness concepts--as in the past seemed like a good place to begin. As was mentioned, most uses of fitness test results do not inform instruction, or give helpful feedback to students and parents, and are probably developmentally inappropriate if they embarrass children. We can do better. Although designing new assessments may be difficult--at least we can begin.

Given the questions of validity and reliability, as well as children and teacher dissatisfaction with traditional fitness tests, along with the new guidelines for physical activity, the recent NASPE standards, and the physically educated person document, coupled with what we know about assessment as a tool for student learning are all reasons for developing an alternative assessment.

Since there is significant evidence that knowing how to be physically active may influence physical activity levels, children's understanding of how to apply physical activity in a personally unique way becomes a meaningful exit outcome. Therefore, three alternative assessments were developed to engage children in active learning and construction of their own knowledge of physical activity.

Specific details of task development can be found in Chapter 3. Furthermore, Chapters 4, 5, and 6 detail each of the three assessment tasks in the following order: Homer Booklet Assessment Task, Video Tape Advertisement Assessment Task, and Perceived Intensity Level Assessment Task.

CHAPTER 3

Methodology

The purpose of this study was to design, pilot, and evaluate three alternative assessment prototypes for fourth and fifth grade children to measure their understanding of physical activity guidelines as contained in the Surgeon General's Report (USDHHS, 1996). Understanding is defined as the ability to apply facts, concepts, and skills, appropriately in new situations (Gardner, 1991). Again, this definition implies synthesis of information and application level understanding--not simple recall. Children applied their understanding of the physical activity guidelines by creating a Booklet for Homer, designing a Video Tape Advertisement for children at another school, and coding their perceived intensity level during the Perceived Intensity Level Assessment.

Four research questions guided the evaluation of all three prototype assessments for this study. They were: 1. Do students who have been taught the physical activity guidelines score differently on the alternative assessment prototype than those who have not been taught the guidelines? 2. Do content experts agree that scores can be used to describe what students have learned (content validity). 3. Do students find the assessment tasks worthwhile, enjoyable, and meaningful? 4. Is the alternative assessment prototype "feasible" for a teacher to use in a regular physical education setting?

Caveat

It will be important to remember the focus of this study was to evaluate the quality and appropriateness of the assessment, not the levels of student performance. In this study, scores for two groups--"experienced" and "inexperienced" with instruction--however, were compared at the end of each

assessment. The National Assessment of Educational Progress (NAEP) used a similar focus in its initial pilot of "Higher-Order Thinking Skills Assessment Techniques in Science and Mathematics" (Blumberg, et al., 1986), but did not compare group scores. NAEP followed a formative evaluation model of designing and refining the tasks through a panel of content experts and piloting with small groups and individual children in numerous settings. A final review and critique of the pilot-tested tasks, review of the data, and suggestions for task revision occurred with a panel of five members, all of whom had participated on one or both of the earlier panels.

This study followed a similar formative evaluation model. It is a first step in the process of designing alternative assessments that measure children's understanding of physical activity and their ability to apply that knowledge. Hopefully, this study will serve as a foundation for researchers and teachers to develop other alternative, perhaps more authentic, assessments that measure physical activity, instead of physical fitness for young children.

This chapter contains the setting, participants, procedures for test development, task development process, data collection, data analysis, and data and researcher trustworthiness sections.

Setting

This study took place at Primrose (pseudonyms used for school, all teachers and students) a small rural Pre/K-5 elementary school located in the Southeastern United States. Approximately 230 students, predominantly Caucasian (90%), considered middle to lower middle Socio-Economic Status (SES), attended the school. Class sizes averaged 16 students in the upper grades with larger numbers in the lower grades.

Kindergarten through fifth grade children attended physical education (PE) classes two days per week for thirty minutes at a time. All instruction and assessments took place during regular PE class time. During this study, all PE classes took place in a rectangular multipurpose room approximately forty-five (45) to fifty (50) feet in length and thirty-five (35) feet in width. This room also served as the cafeteria during breakfast before school, and lunch during school. The stage is located at the south end. The cafeteria kitchen is located on the opposite wall with two doors. At all times, unless they are down for use, long fold-up tables on wheels are stored at the kitchen end between the two doors. Large windows cover the long western side of the multipurpose room and look out onto a pasture where cows were sometimes seen grazing. Opposite the windows was a solid wall, with one shortened, permanently attached basketball hoop. Laminated fitness posters, fitness scores, and other motivational posters were taped to this wall. Classroom rules were located just inside the front double doors where children entered and lined up to exit from class.

Patty was the physical education teacher in this study. She had eleven years of teaching experience in elementary physical education. During this study, Patty was also enrolled in a nearby university completing her final courses to become certified as a classroom teacher. Patty teaches two full days and two half days at Primrose. The rest of the week she teaches physical education at another school in the same district.

Her office was located in a corner behind the stage curtains near an outside exit of the building. There was a teacher's desk, a four-drawer file cabinet, and bookshelves to store supplies. Patty provided a large double student desk and chair for the researcher during the study. The researcher was able to store and keep her things there during the time she was at the school. Patty stores a lot of the small PE equipment (balls of various size, jump ropes, hula hoops, carpet squares,

cones, bowling pins, etc.) behind the curtains on the stage. She also has larger equipment and ball carts stored in a storage shed outside.

A small teacher workroom near the multipurpose room was used to conduct all interviews. The room was small, long, and narrow and was located above the boiler room which made it noticeably warm most days. Three vending machines and one large metal bulletin board paper holder on wheels graced one wall. The opposite wall had floor to ceiling cupboards and shelves for various paper and teacher supplies. A paper cutter and other such equipment was attached to the shelf of the cupboards. A time clock that made a steady rhythmical clicking sound hung on the wall near the door. No desks or chairs were located in this room. The floor was carpeted and there was a short stool used by teachers to grasp supplies stored higher than their reach. The tape recorder was set on this bench during all interviews. All students and the researcher sat on the floor during interviews. Clipboards were used anytime children had to write, as in times when they were reviewing what they had written on their self reflections, or when they were rating their group's performance on the Video Tape Advertisement Assessment.

The researcher was in the setting for fourth and fifth grade PE classes 2 days per week from early January until the end of March. During interviews, the researcher was at the school 3 or 4 days. She entered the setting and tried to get to know as many student names as possible prior to the beginning of the actual study. The first official week was spent distributing informed consents (see Appendix A for Informed Consent). Copies of class rosters were used to keep records of students returning signed (by both a parent/guardian and the student) permissions to participate in the study.

Participants

Students

Two fourth grade classes (N = 33) and two fifth grade classes (N = 32) participated in this study. Two classes, one fourth (N = 17) and one fifth grade class (N= 16) received instruction for how to be healthy through physical activity for a lifetime according to the 1996 Surgeon General's Report. These groups were called "experienced" groups, as in they experienced the curriculum (N = 33) designed for this study. Two classes, one fourth (N = 16) and one fifth (N = 16) did not receive any instructional content, yet participated in the assessments. These groups were called "inexperienced" groups for the study (N = 32).

Those students not returning informed consents (N = 14) signed by both a parent/guardian and the student, or those returning informed consents but indicating they did not want to participate (N = 1) still participated in all instruction and assessments as part of their regular physical education class. (see Appendix A for copies of Informed Consents).

The teacher decided these were important tasks and concepts, and that designing separate instruction for non-participants was not appropriate. Although unconventional, informed consents indicated that student work would not be used for those not participating in the study, but all students would participate in the instruction as part of regular class instruction. Therefore, non-participant work was not included in data analysis for the two product assessments. Due to absences, different numbers of students participated in each of the three assessment tasks. Results of each assessment are contained in Chapters 4 (Homer Booklet), 5 (Video Tape Advertisement), and 6 (Perceived Intensity Level Assessment).

Review Team, Content Experts, and Raters

Two Content Expert Teams served as consultants during this study. They were called the Review Team and Content Expert Team. The Review Team

included the teacher, researcher, and another physical education teacher considered to be an expert at the national level with alternative assessment and young children. The Review Team worked together throughout the process of development, refinement, and evaluation of the three assessment tasks.

The second team was the Content Expert Team. This team was comprised of three elementary physical education teachers pre-selected by the researcher as experienced in assessment due to their use of some form of alternative assessments with children in their physical education programs. The Content Expert Team was brought together to critique and judge content validity of the three assessment tasks after the results were obtained and after the researcher had left the setting. (See Phase 6 for details).

The two adult raters for all the assessments were the teacher and researcher. Scores for the Homer Booklet Task were scored independently at separate times. The teacher and researcher viewed the Video Tape Advertisement Assessments at the same time, but scored them independently. The researcher and teacher coded students during the Perceived Intensity Level Assessment Task independently, but at the same time intervals.

Procedures for Test Development

Although alternative assessment implies new strategies for looking at educational outcomes, the process for developing these assessments is based on decades of measurement research. Developers of high-quality tests, be they norm-referenced, criterion-referenced, or performance-based tests, adhere to steps one through six with certain variations (Herman, Aschbacher, & Winters, 1992).

Steps for Test Development

The six steps they recommended are: "1) Specify the nature of the skills and accomplishments students are to develop. 2) Specify illustrative tasks that would

require students to demonstrate these skills and accomplishments. 3) Specify the criteria and standards for judging student performance on the task. 4) Develop a reliable rating process. 5) Gather evidence of validity to show what kinds of inferences can be made from the assessment. 6) Use test results to refine assessment and improve curriculum instruction; provide feedback to students, parents, and the community" (p. 8).

1. Specify the nature of the skills and accomplishments students are to develop. Understanding of the four physical activity guidelines included:

- 1) a person should be physically active most days of the week;
- 2) at a moderate intensity level;
- 3) for thirty minutes of accumulated activity time.
- 4) Select enjoyable activities that are "fun for you".

2. Specify illustrative tasks that would require students to demonstrate these skills and accomplishments.

- 1) Homer Booklet Assessment Task (Chapter 4)
- 2) Video Tape Advertisement Assessment Task (Chapter 5)
- 3) Perceived Intensity Level Task (Chapter 6)

3. Specify the criteria and standards for judging student performance on the task.

- 1) Holistic Rubric (see Chapter 4, Figure 3).
- 2) Analytic Rubric with two criteria--content and style (see Chapter 5, Figure 9).
- 3) Coding intensity level (percent agreement between adult and child).

4. Develop a reliable rating process.

- 1) Measured inter-rater reliability (two adult raters) for each task.

5. Gather evidence of validity to show what kinds of inferences can be made from the assessment.

- 1) Review Team followed systematic steps throughout the process. The Content Expert Team evaluated content validity based on thirteen questions posed by Herman, Aschbacher, and Winters (1992).

6. Use test results to refine assessment and improve curriculum instruction; provide feedback to students, parents, and the community.

- 1) All three assessments were refined after each administration.
- 2) Teacher interview revealed how scores could be used to show evidence of learning in physical education.

All three alternative assessments for this study were designed following the six systematic steps described by Herman, Aschbacher, and Winters (1992). The initial Homer Booklet Assessment Task can be found in Appendix B. A revised version can be found in Appendix C. The final version can be found in Chapter 4, Figure 3. A more detailed section on test development can be found in Appendix D.

Task Development Process

Three performance assessments were developed for this study. Performance assessments can be: a) constructed responses such as short answer, journals, lab results; b) products such as booklets, video or audio tapes, art projects; c) performances such as hands-on science labs, motor skills, physical fitness tests, gymnastics or dance sequences; or d) process focused such as think aloud, interviews, or conferences (McTighe & Ferrara, 1994). For this study, two of the assessments developed were considered product assessments. See Chapter 5, Figure 7 for details of the Video Tape Advertisement Assessment Task. See Chapter 4, Figure 3 for details of the final Homer Booklet Assessment Task.

Appendix B contains the initial Homer Booklet Task and rubric. The Perceived Intensity Level Task was considered a performance assessment and details are contained in Chapter 6. The design, pilot, and evaluation of the three assessment tasks were conducted in primarily 7 Phases.

Phases of the Study

The process was similar, but each assessment was developed a little differently, and is, therefore, described separately. The following section describes each assessment task separately according to the seven phases listed below.

Phase 1: Develop and pilot with small groups

Phase 2: Revise with teacher and researcher

Phase 3: Administer with intact class first grade level

Phase 4: Revise, if needed, with review team

Phase 5: Administer with intact class other grade level

Phase 6: Review Team meets to evaluate all tasks and make any suggestions for further revisions.

Phase 7: Content Experts evaluate content validity and make any suggestions for further revisions.

Homer Booklet Assessment Task

Phase 1: Develop and pilot with small groups.

The Homer Booklet Task was developed and piloted with small groups of two to four children at a pilot school called Wilson Elementary. This is a different school from where the actual study took place.

Phase 2: Revise with teacher and researcher.

Students and the physical education teacher helped modify the booklet design to make it more "child friendly." For example, students suggested folding

the paper to make the actual Booklet look like a booklet. The teacher suggested putting borders around the blank pages to give the students boundaries within which to write.

Phase 3: Administer with intact class first grade level.

The Homer Booklet Task was administered to both fourth and fifth grade intact classes who had received instruction at Wilson Elementary. At Primrose Elementary, the Homer Booklet Assessment was first administered to fifth grade "experienced" and "inexperienced" groups.

Phase 4: Revise, if needed, with Review Team.

The Review Team met and suggested a revision of the rubric to focus primarily on the content (guidelines), versus the quality of the student work (see Appendix B and C for the initial and revised Homer Booklet Rubric and see Chapter 4, Figure 3 for Final Homer Rubric).

Phase 5: Administer with intact class other grade level.

Changes were made to the scoring rubric to emphasize the number of guidelines included in the booklet, versus the quality of the student work. Next, the Homer Booklet Task was administered to fourth grade.

Phase 6: Review Team meets to evaluate all tasks and make any suggestions for further revisions.

The Review Team met and agreed that student scores could be judged more accurately and less subjectively using the final scoring rubric. Although some rubric developers might consider this a simplistic rubric, it was agreed to keep it as is (see the final rubric for the Homer Task in Chapter 4, Figure 3).

Phase 7: Content Experts met to evaluate content validity and make any suggestions for further revisions.

Specifically, during the final phase of evaluation, a panel of Content Experts convened to evaluate and make professional judgments about the three assessment

tasks (see Content Expert Training Agenda in Appendix E). Experts were offered substitute pay for their full day of missed school. One district paid for their two teachers, as they considered the session to be an in-service and valuable professional training. The other two districts graciously allowed their three teachers to take professional leave days, and the researcher paid for their substitutes for the day. The combined teaching experience for all six content experts totaled 84 years.

The Review Team and Content Experts were present during the evaluation. Cara, from the original Review Team, served as moderator to keep us on task. Patty, the teacher, took notes. Both contributed to answering questions pertinent to their roles, insights, and perceptions in the development process.

The three teachers from the Content Expert Team asked questions about the instruction, process of development, and administration of assessments. Cara suggested that the team consider going back to the original rubric in order to distinguish quality, and not just the number of responses.

Validity and Reliability

Validity was defined as an assessment that measures what it says it measures. Content Experts were told they would answer thirteen questions pertaining to content validity that afternoon (See Chapter 4, Figure 5 for an example of the thirteen validity questions). It was explained that no value judgments (validity) could be made without valid reliability measures.

Reliability was defined as consistent scoring between two raters--or inter-rater reliability. Inter-rater reliability was calculated by counting the number of possible perfect score matches, versus the number of matches made by both raters at the perfect match level. For example, there were forty-four possible scores for the Homer Booklet Task. Both raters scored the children at the exact same rubric score--forty-three (43) of the forty-four (44) possible times. Forty-three divided by

forty-four equals ninety-eight percent. Inter-rater reliability percentages were given for each assessment. Content Experts were told that the Homer Booklet Task inter-rater reliability scores were in acceptable ranges.

Finally, the Content Experts rated each of the three assessments according to content validity questions previously written by the researcher (see each relevant chapter for content validity rubric). These questions were based on criteria recommended by Herman, Aschbacher, and Winters (1992).

Each assessment was set up at a station. All paper work pertinent to that assessment task, including all the tasks and rubrics from the beginning through all revisions to the end, as well as all student responses, were set out to be reviewed. Any and all questions could be asked to help clarify the process, task, or rubric for them. The three Content Experts rotated to each station. One of the original Review Team members (researcher, teacher, teacher-assessment expert) was assigned to each station to further clarify or answer any questions (see Appendix E for Content Expert Agenda).

Video Tape Assessment Task

Phases of the Study: The second assessment task followed a similar format to the development of the first one.

Phase 1: Develop and pilot with small groups.

Three groups of two or three children at Primrose Elementary School were selected to participate in the development and pilot of the all the assessment tasks. Initially children were taught the following four guidelines from the Surgeon General's Report. The guidelines were: 1) A person should be physically active for thirty minutes per day, 2) at a moderate intensity level, 3) most or all days of the week, and 4) select activities you enjoy that are "fun for you" (see Appendix F for teacher instruction).

Three primary recommendations are included in the Surgeon General's Report for all persons over the age of two. They are the first three guidelines from the previous paragraph. Enjoyment was included because it is the major reason, or determining factor, children engage in physical activity. Both the researcher and teacher felt these four guidelines were important for youngsters to learn, so they might know "how to" be physically active to be healthy in contrast to physically fit.

Youngsters were shown student examples of the Homer Booklet Assessment Task and told it was one possibility adults could think of to ask children to "show off" what they know and can do. The researcher explained that she wanted their ideas. Thus, children were asked: "What would be a 'fun way' for you to 'show off' what you know?" Children suggested creating a video tape movie that could "tell and show" children how to follow guidelines to be physically active for a lifetime. The task was then developed and completed with these three pilot groups.

Phase 2: Revise with teacher and researcher.

The Video Tape Assessment Task followed a written format similar to that developed and used by Educators in Connecticut (1996). See Chapter 5, Figure 7 for written details of the Video Tape Advertisement Assessment Task.

The Review Team suggested minor editorial changes to help clarify the procedures section from the student video task description. (See Chapter 5, Figure 7). At this point, it was decided that the rubric would be further developed after student examples were available.

Phase 3: Administer with intact class first grade level.

Fourth grade classes at Primrose Elementary completed the Video Tape Advertisement Assessment Task.

Phase 4: Revise, if needed, with Review Team.

In Phase 4, the rubric was revised. The researcher suggested that an analytic rubric might work better for this assessment task. The Review Team agreed. It was decided that there would be a content and style score. Scores would be weighted so content would count more than style. In other words, a group doing a video presentation, without the information, would not score as high as a group who included the guidelines, but lacked style. These changes were made, and video assessments were scored by the teacher and a second adult rater. (see Appendix G for Video Tape Advertisement Assessment--Weighted Version).

Phase 5: Administer with intact class other grade level.

The video task was administered to fifth grade classes at Primrose using the new scoring rubric. Students completed the design and practice in two days. It took two more days for all groups to be taped.

Phase 6: Review Team met to evaluate all tasks and make any suggestions for further revisions.

The researcher suggested the two scores (content and style) should stand alone and not be combined as a weighted score. The Review Team agreed that showing both a content and style score would better represent the students understanding or knowledge. (See Chapter 5, Figure 8 and 9).

Phase 7: Content Experts met to evaluate content validity and make any suggestions for further revisions.

The same procedures as described in the Homer Booklet Assessment Task occurred in Phase 7 of the Video Tape Advertisement Assessment Task. Inter-rater reliability was determined by the same method. The number of perfect score matches between the two raters were summed and divided by the total number of possible matches. Inter-rater reliability for the Video Tape Advertisement Assessment Task was 94%.

Perceived Intensity Level Assessment

The third assessment task followed a similar format to the development of the first and second task.

Phase 1: Develop and pilot with small groups.

Groups of two to four students at the pilot school--Wilson--were asked by the researcher what types of active stations they enjoyed. The researcher asked students to describe their favorite active stations. From this list, the teacher and researcher created five stations for the Perceived Intensity Level Assessment Task. The stations that were designed allowed the students to be active at any intensity level. (See Chapter 6, Figure 15).

Phase 2: Revise with teacher and researcher.

The task was written following the same format as described in Phase 2 of the Video Tape Advertisement Assessment Task. See chapter 5, Figure 7 for the format. A student coding system was created. Figure 1 and 2 located in Appendix H detail the first coding instruments.

Phase 3: Administer with intact class first grade level.

At Wilson, the pilot school, two intact fifth grade classes experienced the Perceived Intensity Level Assessment. At Primrose, two intact fourth grade classes, one "experienced" with instruction, one "inexperienced" receiving no instruction, performed and completed the Perceived Intensity Level Assessment Task (see Appendix H--Student Designed Station Task1). The teacher interviewed students and discovered the coding could be changed to make it easier for the youngsters.

The teacher and researcher decided the task purpose did not match well with the task. Therefore, it was decided that the researcher and teacher would create the stations and students would only code, not create. See Chapter 6 for further details.

Phase 4: Revise, if needed, with Review Team.

Coding changes were made. See Appendix H--Figure 1 and Figure 2 for details.

Phase 5: Administer with intact class other grade level.

The Teacher Designed Station Assessment was then administered to the same fifth grade groups as had completed the Student Designed Station Task (see Appendix I for details of the Teacher Designed Tasks).

Phase 6: Review Team met to evaluate all tasks and make any suggestions for further revisions.

The circle coding system was further revised to include a line on the circle on which to record the partner's score (see Chapter 6, Figure 14 for the Final Circle Task Perceived Intensity Level Student Coding Instrument).

Phase 5 again: Administer with intact class other grade level.

The "Circle Station Task", now called Perceived Intensity Level Assessment, was administered to fourth grade intact classes. Students participated in the self-assessment on day one of the assessment, and self and partner assessment on day two. See Chapter 6 for further details of the task and coding instrument (Figure 13, 14 and 15).

Phase 7: Content Experts met to evaluate content validity and make any suggestions for further revisions.

Content Experts met on October 11. As in the previous two assessment tasks, the same procedures were followed. Inter-rater reliability was calculated as the number of possible matches between two raters and the number of actual perfect matches. Inter-rater reliability scores are reported in Chapter 6 (Content Validity Section).

Data Collection

Data collected during this study included: 1) student reflections; 2) interviews; 3) open-ended question survey; 4) supplemental data including all assessment tasks, rubrics; and documented instruction 5) teacher interview; 6) Content Expert's evaluation. The data collection section details each of these sources by its numerical order.

1) Student Reflections: Students responded in a self-reflection in written format to each assessment task. See Chapter 4, Figure 6 for an example of the student self-reflection. Some students were also interviewed after each assessment, and many were interviewed after all assessments were administered.

2) Interviews: Students were interviewed during the development and pilot of the assessment tasks to make necessary and appropriate changes according to their feedback. Students were also selected and interviewed following final administration of all assessments to gain deeper insights into their responses on self-reflections. Student interview data was used for triangulation purposes. All students were interviewed using an interview guide. (see Appendix J for Interview Guide questions for students).

Students were interviewed in their assessment groups by the researcher for the Video Tape Advertisement Assessment. For the other assessments, two students with similar results and student reflection responses were interviewed together. This method had been used successfully in previous studies (Hopple, 1994; Manross, 1994; Sanders, 1993).

3) Questionnaire "Your Favorite Assessment": After all assessment tasks were completed, a final questionnaire of open-ended questions was filled out by each student. This was decided post facto because two of the assessments measured the same content. Therefore, the researcher believed that discovering which task

students preferred, when given a choice, would be interesting information (see Appendix K for "Your Favorite Assessment").

4) Supplemental Data: Supplemental data includes the three assessment tasks, rubrics, scoring guidelines, and instructional materials. All tasks and rubrics can be found in each relevant chapter (see Appendix F for Teacher Instruction).

5) Teacher Interview: The teacher interview was conducted to gain Patty's insights into the feasibility of each assessment, and the process of the development, pilot, and evaluation of the assessments for fourth and fifth grade students in a regular physical education setting.

6) Content Experts: Three Content Experts in elementary physical education, as well as the Review Team (researcher, teacher, and one other Content Expert), participated in the final phase of the evaluation process. An analytic rubric was created with 13 questions relating to content validity. See Chapter 4, Figure 5 for an example the 13 content validity questions. The three Content Experts evaluated content validity for the three alternative assessments based on those thirteen questions.

Data Analysis

This study describes the development of the three alternative assessments. Qualitative data analysis and descriptive statistics were used to describe and interpret meaning. All four research questions were answered for each of the three alternative assessments and data was analyzed according to each of the four research questions. The following section details how data was analyzed for each research question and which data sources were used to answer each question.

Research Question # 1

Do students who were taught the guidelines score differently than students who were not taught? Group means and percentages were calculated from rubric score results. The teacher and a second adult rater independently rated the booklets. An average of the two scores was calculated by adding the teacher and other rubric scores together and dividing by two. However, only percentages of students matching adult raters were calculated for the Perceived Intensity Level Assessment Task.

Research Question # 2

Do Content Experts agree that scores can be used to describe what students have learned (content validity)? Content Experts were brought together to gain their insights and ask for their judgments about the content validity of each assessment task. Content Experts spent the morning in a training session (see Appendix E for Agenda) and the afternoon evaluating each assessment task according to the validity scoring rubric.

Each of the three assessment tasks were displayed at one table. All student scores and responses for the particular assessment task were available for perusal at each table. Content Experts rotated to each table to evaluate each assessment task one-at-a-time. The researcher and others on the Review Team were available to answer any questions about the process, instruction, and student scores. Mean scores for content validity were calculated by summing the total score for each Content Expert and dividing by thirteen, the number of questions. The results and discussion are included in each relevant chapter for each assessment task. Teacher interviews and Content Expert comments were included to triangulate validity data.

Research Question # 3

Do students find the assessment tasks worthwhile, enjoyable, and meaningful? Student self-reflections were analyzed first by simple number counts according to overall enjoyment for all three assessment tasks. Student self-reflections contained seven (7) questions related to student perceived enjoyment, difficulty, and importance of physical activity knowledge, and was the primary data source analyzed for Research Question # 3. The number of yes and no responses were counted and reported. The percentage of yes responses within each intact class and overall percentages for all students, for example, N = 44 for the Homer Booklet Task were calculated for all three assessment tasks.

Responses were first compiled according to the specific question on the self-reflection. For example, student reflection Question 1 asked "Overall, did you find this assessment to be enjoyable? Why? Why not?" (See Chapter 4, Figure 6 for an example of student reflection questions). Most students gave at least one reason why, or why not. Some students gave more than one reason. Reasons were analyzed according to a constant comparative method as described by Bogdan and Biklen (1982).

All student written responses were first copied. Responses containing more than one reason were copied that many times. Next, those responses were pasted to poster boards according to common broad themes (Bogdan & Biklen, 1982) related to their enjoyment and the meaningfulness questions. A similar process was used successfully in previous studies (Graham, Hopple, Manross, & Sitzman, 1993; Hopple, 1994; Manross, 1994). In an attempt to prevent any researcher bias, broad themes were confirmed by a second researcher.

Interview data was a second data source for triangulation of the third research question regarding student meaningfulness or enjoyment. All interviews were first transcribed and typed into the computer using Word 5.1. Interviews were read many times. Interview statements confirming student's written

responses in their student self-reflections, were color coded with post-it flag tags and included under the same student self-reflection questions on the poster board. (See student meaningfulness section in each assessment chapter for student examples). Interviews, therefore, provided triangulation in the data to confirm student responses and insights, as well as to provide deeper and broader understanding of student's perceptions and insights of the assessment tasks.

Research Question # 4

Is the alternative assessment prototype "feasible" for a teacher to use in a regular physical education setting? The teacher interview was transcribed and Patty's perceptions about the feasibility of the assessment were drawn from this interview. Content Expert judgments provided triangulation for feasibility data. When available, written statements by Content Expert's about feasibility were used to report their insights. A third source used to judge feasibility was the ease of response by students as perceived by the teacher and researcher. This information is contained in the researcher's field notes.

Final Survey

Final surveys (see Appendix K) were used to discover which assessment students preferred the most. Simple number counts were used to determine how many students enjoyed which assessment, and which assessment was enjoyed by what percentage of the students in this study. This data is included in Chapter 7.

Data and Researcher Trustworthiness

Steps were taken to ensure confidence in the data collection and analyses of the alternative assessments. This researcher was intimately involved in designing, piloting, and evaluating the three assessment instruments in collaboration with the teacher and youngsters. Therefore, the researcher was closely connected to the data. It will be up to the reader to make judgments about the trustworthiness of the

data and researcher. This researcher was a very active participant throughout all phases of the process. She maintained an active interest in the process, and was present in the research setting for a prolonged period of time (Locke, 1989).

Another way to establish confidence in the data collection and analyses is through the process of cross-checking data. In this study, student reflections and interviews were cross-checked to confirm or disconfirm consistency in statements. These steps help prevent erroneous conclusions. Data sources were checked (Locke, 1989) with students, the teacher, and Content Experts, by the researcher to verify comments or meanings of written documentation as deemed necessary. Furthermore, as data were collected, the Review Team met to review and critique the raw data. This peer debriefing process served as a means of guarding against personal bias and the influence of the researcher on the study (Locke, 1989).

Summary

This chapter detailed the setting, participants, test development procedures, task development process and steps, data collection, data analysis, and researcher and data trustworthiness related to this study. Chapters 4, 5, and 6 contain the results and discussion for each assessment task. Chapter 4 details the Homer Booklet Assessment. Chapter 5 details the Video Tape Advertisement Assessment. Chapter 6 details the Perceived Intensity Level Assessment. Chapter 7 is the final chapter that includes a final summary and conclusion for all three assessments.

CHAPTER 4

Results & Discussion

Homer Booklet Assessment Task

Chapter 3 detailed the methods and procedures used for developing the three assessment tasks for this study. Chapter 4 details the Homer Booklet Assessment Task and results. The purpose of this study was to design three alternative assessment prototypes. The assessments would measure fourth and fifth grade student's understanding of physical activity as contained in the Surgeon General's Report.

Four research questions guided the evaluation of the assessment tasks. They were: 1. Do students who have been taught the physical activity guidelines score differently on the alternative assessment than those who have not been taught the guidelines? 2. Do content experts agree that scores can be used to describe what students have learned (content validity)? 3. Do students find the assessment task worthwhile, enjoyable, and meaningful? 4. Is the alternative assessment prototype "feasible" for a teacher to administer in a regular physical education setting?

This chapter contains sections including: 1. Homer Booklet Assessment Task Description; 2. Score Results; 3. Content Validity; 4. Student Meaningfulness (enjoyment); 5. Feasibility; 6. Conclusion and Interpretation of Results.

Homer Booklet Assessment Task Description

The Homer Booklet Assessment Task was an assessment, in booklet format, that asked students to write to Homer--a peer--how to be physically active for a lifetime. For the purpose of this study, the Homer Booklet Assessment Task is considered a product assessment under the large umbrella term of performance

assessments according to the National Education Association (McTighe & Ferrara, 1994).

The Homer Booklet Assessment Task was designed to assess how well the students understood the revised guidelines constituting beneficial physical activity (USDHHS, 1996). The four criteria are: 1) a person should be physically active most days of the week; 2) at a moderate intensity level; 3) for thirty minutes of accumulated activity time. 4) Children were also taught to select activities that are "fun for you". These criteria were taught to the experienced group. Due to some absences only forty-four students participated in this assessment. Figure 3 displays the Homer Booklet Assessment Task and rubric as read to the children. See Appendix L for completed student examples of the Homer Booklet Assessment Task.

Homer Score Results

This section highlights rubric scores by group means, the number of students and percentages of students in specific groups scoring at specific score points. This section helps answer research question # 1: Do students who have been taught the physical activity guidelines score differently on the alternative assessment prototype than those who have not been taught the guidelines? Figure 4 displays the mean scores in a histogram for the Homer Assessment Task.

Final Homer Booklet Assessment Task

Homer does not know about the new Surgeon General's Report. He does not know this information because it is so new. This report explains how to gain health benefits through a lifetime of physical activity. In this class, you have learned how to be physically active to meet the standards for health benefits. You can help Homer learn this information by designing a booklet that he can read. The booklet should include all of the information Homer needs to be physically active for a lifetime. The booklet should also be interesting and enjoyable for him to read. Include drawings and art work to help Homer understand.

RUBRIC

Level 4: Surgeon General

- The student clearly and accurately explains all the guidelines needed to obtain health benefits.
- The student clearly explains to Homer how to choose activities that will help Homer continue the activities.
- The booklet should be interesting and enjoyable to read.
- The booklet should be publishable.

Level 3: Surgeon General's Aide

- The student leaves out one guideline necessary to obtain health benefits.
- The booklet may or may not be interesting and enjoyable for someone else to read.
- The booklet may or may not be publishable.

Level 2: Apprentice Aide

- The student clearly and completely explains half of the guidelines, but, two important parts are missing. Therefore, Homer will not quite know how to be physically active for a lifetime to obtain health benefits.
- The booklet may, or may not be interesting and enjoyable to read.
- The booklet may or may not be publishable.

Level 1: Novice Aide

- Less than half of the guidelines are included, therefore, Homer will not know how to be physically active for a lifetime to obtain health benefits.
- The booklet may or may not be publishable and enjoyable to read.

Level 0:

- The student chooses not to do the booklet.

Figure 3. Homer Booklet Assessment Task (Final version)

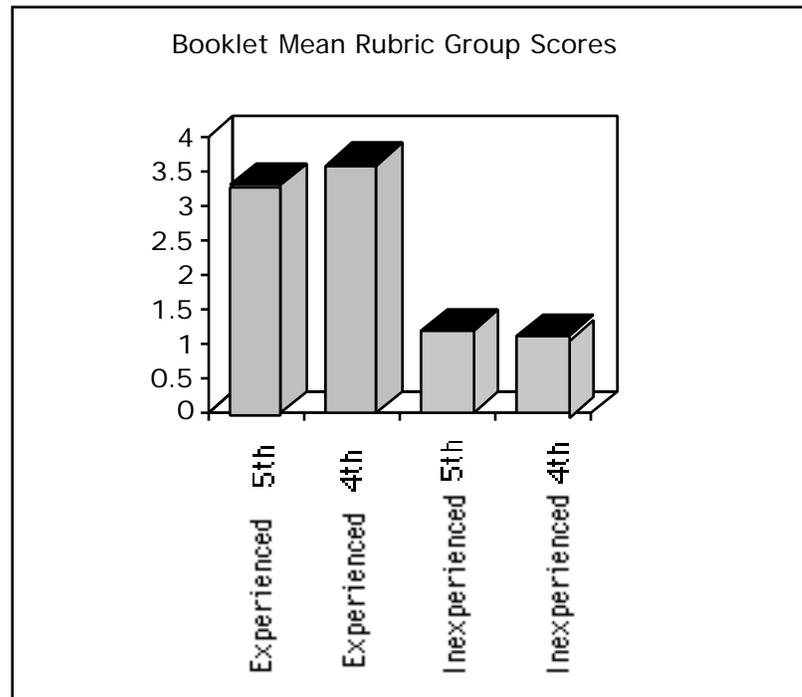


Figure 4. Fifth grade mean rubric scores were 3.3 for the "experienced" group and 1.2 for the "inexperienced" group. Fourth grade mean rubric scores were 3.6 for the "experienced" group and 1.1 for the "inexperienced" group. The total was out of a possible rubric score of 4. Table 1 displays the percentage of students in fifth grade scoring at each score point.

Table 1

Fifth Grade Homer Booklet Rubric Scores & Percentages

Experienced 5th Grade			Inexperienced 5th		
Rubric Score	# Students	Percentage	Rubric Score	# students	Percentage
1	1	9%	1	9	82%
2	0	0%	2	2	18%
3	5	45.5%	3	0	0%
4	5	45.5%	4	0	0%

Table 2

Fourth Grade Homer Booklet Rubric Scores & Percentages

Experienced 4th Grade			Inexperienced 4th		
Rubric Score	# students	Percentage	Rubric Score	# students	Percentage
1	0	0%	1	9	90%
2	2	14%	2	1	10%
3	5	36%	3	0	0%
4	7	50%	4	0	0%

Ninety-one percent of fifth grade students in the "experienced group scored 3 or 4 on the rubric with one student scoring a 1. In contrast, eighty-two percent of fifth grade "inexperienced" students were only able to accurately identify one or none of the four guidelines. Table 2 displays the percentage of fourth grade students scoring at each rubric score point.

Fifty percent of fourth grade "experienced" students accurately identified all four guidelines. Five (5) more students scored 3's indicating they recalled only three guidelines. Conversely, nine of ten "inexperienced" fourth grade students recalled only one or none of the guidelines (score of 1). Thus, students who were taught the guidelines scored higher than those not taught the guidelines indicating that the instrument discriminated between students who were taught and not taught. Interview data for youngsters scoring 2's on the rubric from the "inexperienced" groups revealed that they had some previous knowledge.

Homer Validity Question/Content ExpertsResults and Discussion

Validity and reliability concerns are important and critical to any test or assessment. In standardized tests, however, validity often gets sacrificed for

reliability. In fact, tests generally only measure what is easy to measure, and validity typically gets sacrificed for reliability of results, efficiency of scores, and ranking (Wolf, et al., 1991). For alternative and authentic assessments validity becomes very important. Because alternative and authentic assessments take place in the context of learning, they are not necessarily as standardized. That's the whole point (Wiggins, 1993). There should be opportunity for synthesis of information and opportunity for diverse responses for individual assessments or group assessments. Additionally, professional judgments are made, so rater training and consistency in scoring become important (Herman, Aschbacher, & Winters, 1992).

The validity section highlights important reliability, and particularly, validity questions. Inter-rater reliability scores for all assessments were within acceptable standards and were reported to the Content Experts. Content Experts were told that an assessment or test can not be considered valid if scores are not reliable. Inter-rater reliability for the Homer Booklet Assessment Task was 98% for two raters at the perfect match level. The two adult raters independently scored the rubrics. The researcher calculated the number of perfect agreement matches compared with the number of possible matches to obtain the percentage agreement between the two raters. Content validity was evaluated by the teacher's perceptions and thirteen questions answered by the Content Expert Team. (See Figure 5).

One important question for content validity is: Do scores accurately discriminate between those students who were taught the content and those who were not? The researcher asked Patty, the teacher, this very important question.

Patty: Oh yeah. I mean, like I said, it definitely showed how much of it they understood and what they didn't understand. Most of the learning group in fifth grade had mostly 4's or 3's...which they had all the information... and the 3's they lacked one... and on several occasions it was

just the simple fact of being active by doing something you like--which is the one, if I remember correctly, that was left out the most. So the hard core facts were there...

Researcher: Right.

Patty: ...So you could take, especially the booklet, we've been going over these 4 things, these kids wrote a letter, or wrote directions to this fictitious person about the concept of being healthy for a lifetime and a principal would definitely see--there it is.

Another content validity concern is the meaningfulness and worthiness of the content. Is the content important--an essence of the subject matter--or is it frivolous information not worthy of the time? Patty was asked this question. She said:

Patty: I think it rates up there with you're trying to teach your children about nutrition and you're trying to teach them how to make healthy choices in their life as far as what they eat...

Researcher: As a parent you mean?

Patty: As a teacher, parent...providing kids with knowledge however, whatever direction you're coming from...you want to teach them that they have choices about what they eat and what things provide the good parts for their growing and their development and being physically active--physical education, to me being physical and being a physical educator that, in my mind, I don't see a lot of difference.

Researcher: Okay.

Patty: Cause physical education to me is teaching a child about their body and how to make it move and how to keep it healthy. All that is synonymous with me with physical education...which is totally different from physical fitness, which I don't teach physical fitness, I teach physical education.

Researcher: So it would be meaningful content then?

Patty: Yes, it's, it's all the same to me.

Three Content Experts evaluated the Homer Booklet Assessment Task, in the attempt to determine if, or how, scores can be used to make judgments about student learning. All student data, including scores and student responses, were

available. All teacher and student task directions and scoring rubrics, including all changes from beginning to end, were also made available for Content Experts to examine. Three Content Experts were asked thirteen questions for each assessment. Questions were modified from criteria set by Herman, Aschbacher, and Winters (1992). They state that teachers can be fairly confident that an assessment has content validity if teachers can answer yes to most of the questions. Raters were asked to judge the assessment according to the questions and rubric found in Figure 5.

Content Expert, Josh, scored question number 10 as a one. Content Expert, Calvin, scored question number 13 as a one. A score of 1 means a revision may be necessary, or depth, or quality is lacking. Content experts judged all but the previous two questions on the Homer Booklet Assessment as "clearly meets criteria without hesitation or reservation". In fact, the average overall rubric score for all the questions by the three outside Content Experts was 1.93 out of a possible score of 2. Therefore, the Content Experts agreed that the Homer assessment design satisfactorily met the criteria for content validity.

It is important to point out to the reader, however, that extreme caution should be used when interpreting these results. These scores are from small groups of youngsters from one physical education setting or program, one teacher, and three content experts. Therefore, the author would caution against generalizing to other programs and populations. These assessments and rubrics would need to be further piloted and evaluated in many diverse settings with lots of youngsters before true judgments could be made about them.

Content Validity Rubric (Homer Booklet Assessment Task)

<p>Rubric for rating:</p> <p>2 Yes, clearly meets criteria without hesitation or reservation.</p> <p>1 Yes--but depth or quality is lacking, and revisions may be necessary.</p> <p>0 Cannot answer yes to the question, therefore does not meet the criteria to the question.</p>
--

Validity Questions: Can the scores be used to describe what students have learned?	Rubric Score
1. Does the content of this assessment reflect important content in physical education curriculum?	
2. Is there a good match between the skills and knowledge intended for this assessment and those emphasized in this PE program?	
3. Is there a clear definition of the assessment goal or purpose so that judgments can be made about the match between knowledge and understanding intended for this assessment	
4. Does the assessment task match the intended instructional goal?	
5. Does this assessment task measure students understanding of physical activity?	
6. Does the assessment task include scoring criteria?	
7. Does the criteria include standards for judging the student performance?	
8. Does the rubric or scoring criteria appear to discriminate between those who understand and those who need further assistance?	
9. Is the task developmentally appropriate for fourth and fifth grade students?	
10. Does the assessment task reflect processes and outcomes suitable for the fourth and fifth grade students?	
11. Did students have sufficient opportunity to learn what is included in the assessment?	
12. Does this assessment task require students to apply their understanding of physical activity in new situations?	
13. Does the scoring rubric accurately describe students understanding of the guidelines for how to be healthy through physical activity for a lifetime?	
Average Rubric Score for the validity of this assessment	

Is this a feasible assessment for regular physical education classes? Why or why not?

Figure 5. Content Validity Rubric (Homer Booklet Assessment Task)

Homer Meaningfulness Question

Results & Discussion

This section highlights the third research question: Do students find the assessment task worthwhile, enjoyable, and meaningful? Wiggins (1993) argues that motivation and the quality of the response increases if the student finds the assessment meaningful. For this study, the word enjoyment was used to measure student's perceived meaningfulness because when students were interviewed it became clear that the word meaningfulness did not mean much to them. Enjoyment, however, was a word children understood and could relate to therefore enjoyment was used to imply meaningfulness. Furthermore, the last student reflection question asked students if they thought the information they were learning was important to know. The fact that students perceived the information as important to know was also interpreted by the researcher as one indicator of meaningfulness to the students.

The word enjoyment was taken from the "Aquarium Problem" from the New Standards Project (1992). In this project, fourth grade students were asked to respond in a self-reflection: 1) What did you enjoy about the task? 2) What did you not like about the task? 3) How is this task like other activities you do in your class? How is it different?

For all three assessment tasks in this study, students responded in a student self-reflection to seven questions. See Figure 6 for student self-reflection used in this study. Table 3 contains the number and percentage of yes and no responses from the first question in the self-reflection that read: "Overall, did you find this task to be enjoyable? Why? Why not?"

You can help us make these learning activities even better. Think about each of the following questions, and write to us what you honestly think. Be as clear as you can (you might want to give us examples of what you mean). You may use the back of this page if you need more room.

1. Overall, did you find this task to be enjoyable? Why? Why not?
2. What did you enjoy about the task?
3. What did you not like about the task?
4. Are there any changes you would like to make to this task? If so, please explain or give an example.
5. Did you find this task (easy, difficult, about average)? Why?
6. Do you think other students your age would enjoy this task? Why, or why not?
7. Do you think it is important to know about physical activity and health? Why? Why not?

Figure 6. Student Self-Reflection and Ideas

Overall, 96% of the students (N = 24) in the "experienced with instruction group" enjoyed the Homer Booklet Assessment Task while 79% of the "inexperienced group" enjoyed the task (N = 15). Most students, whether instructed in the content or not, enjoyed the Homer Booklet Assessment Task. It appears that there may be a difference in the "enjoyment" of the task between those receiving and those not receiving instruction.

The majority of students--39 of 44-- enjoyed the Homer Booklet Assessment Task. This is not surprising. Sallis, McKenzie, and Alcaraz (1995) found similar results in curriculum evaluations. They found that elementary school children could evaluate components of a physical education curriculum although they "liked" all of the activities in the curriculum. They also found that students preferred some activities over others.

Table 3
Booklet Assessment Overall Task Enjoyment

Groups	n	<u>Yes</u>	<u>No</u>
5th Experienced	11	10 (91%)	1 (9%)
5th Inexperienced	9	7 (78%)	2 (22%)
4th Experienced	14	14 (100%)	0
4th Inexperienced	10	8 (80%)	2 (20%)
Experienced Total	25	24 (96%)	1 (5%)
Inexperienced Total	19	15 (79%)	4 (22%)
Totals	44	39 (89%)	5 (11%)

*note: 1 student found the task okay and that score was included in the yes response.

Yes = # of students who enjoyed the assessment task

No = # of students who did not enjoy the assessment task

In their own words, what are youngster's reasons for liking or enjoying the assessment tasks? What did these fourth and fifth grade students write in response to questions in the student self-reflection? In interviews, what did children say about their enjoyment of the Homer Booklet Assessment Task?

Meaningfulness (Student Self-Reflection Question # 1 & # 2)

Four broad themes emerged from the data pertaining to meaningfulness or student enjoyment. These student responses come from the first two student self-reflection questions from Figure 6. 1) "Overall, did you find this assessment to be enjoyable? Why? Why not?" 2) "What did you enjoy about the task?" Student responses from these two questions were combined due to similar answers.

Interview data was also used to triangulate student responses. Theme 1: A large group of students found the assessment enjoyable because they like to write, draw, or both write and draw. Theme 2: Students enjoyed thinking about, or being challenged to think about, physical activity as it related to their own health. Theme 3: Another group enjoyed the booklet assessment because they got to help others, write or explain in their own words, or explain to someone else how to be physically active for a lifetime (authenticity). Theme 4: Some students enjoyed the booklet assessment because being or staying healthy is important to them, or became important to them.

Examples in the student's own words detail enjoyment of the Homer Booklet Assessment within the four themes. Quotes are typed in exactly as found in the student reflections including spelling and punctuation. Interview data is also used to support the themes. Students found the assessment enjoyable because...

Theme 1: They like to write, draw or both write and draw.

"Yes Because you can wright a lot and draw" [Kendra BC4]

"Yes, Because you got to write" [Rodney BS5]

Yes because I like to be able to add pitures for visual aide" [Jim BS4]

"Yes I got to drall." [Jenny BS4]

Interview data supports similar statements.

Alison [BC4] said: "Yes...Because I like to draw and write and explain things."

Me: "Just explain or explain to other people?"

Alison [BC4]: "Explain to other people."

Alison's statement also reflects her perception of having a real audience to help--authenticity. Wiggins (1993) states that authenticity, whether real or

perceived, tends to heighten student's motivation level and quality of their work.

In a follow-up interview, Jenny [BS4] elaborated:

I liked doing the booklets too because I like to draw and I liked the way you all put the booklet and, you know, let us like make our own booklet and you didn't like exactly say this is a test and you don't have to study for it and you don't have to do this. Because if it was a test I would have been like this (put hand to chest, eyes got big and she made a noise indicating fear).

Theme 2: They enjoyed thinking about or being challenged to think about physical activity as it related to their own health.

"Yes, because I like thinking about it! [Ira BC4]

"Yes because we got tested on our knolage" [Cameron BS5]

Interview data supports the notion that children where challenged to think while doing the Homer Booklet Assessment Task. In her interview, Ira [BS5] elaborated on her perception of both thinking and authenticity, both themes #2 and #3. She said:

Well, I liked thinking about it cause try thinking of examples, and, you know, I liked writing it to someone else...I think they'd (referring to other children liking this task) like thinking about it because, I mean, you do have to think a little how you'd write someone that's your age.

Theme 3: They felt they were able to help others; write or explain in their own words; or explain to someone else. [Authenticity]

"Yes because I got to do something I liked because I like helping people" [Briana BS5]

"Yes, it is fun writting to an imaginary person" [Angie BC4]

"Yes because you wrote in your own words how to be healthy for a lifetime" [Lee BS5]

"Yes you got to help people" [Jerry BS4]

"Yes because I was giving advise" [Kara BH5]

Interview data supported these children's perceptions of meaningfulness or authenticity. A fifth grade student, Lee, [BS5] said:

Well, you know, when you write it and people tell you a certain way how to do it. Well, this time you could explain, um, in your own words how to do it. Like, if you want to give a lot of examples of it you could do that. Let's say you didn't want to give any examples at all. You could do that. That's what I mean by doing it all in you own words.

Theme 4: Staying or being healthy is important to them or became important to them.

"Yes, because I want to keep my body healthy For a life time. [Scott BC4]

"I found it enjoyable over all because it helped me understand that you need to be healthy for a lifetime. [Ann BS5]

"Yes, because you can be healy and not fat" [Ray BS4]

"Everything because it is important" [Jesse BH5]

In her interview, I asked Ann what she meant by what she had written. She elaborated: "Um, because you really need to know cause if you don't know that then you won't be as healthy as if you did know that." When Scott [BC4] was asked why he said he enjoyed the booklet, and if he thought others would, this was his response:

I like it because I want to keep my body healthy and be physically active for a lifetime. And I don't know about other...if other people want to keep their body healthy...they might keep physically active. If they don't, they might just get disease and die.

Meaningfulness (Student Self-Reflection Question #3)

In this study, students evaluated the Homer Booklet Assessment Task according to their overall enjoyment. Because of such positive results, the researcher also decided to analyze negative and okay responses. Very few students

(N = 5) responded that they did not like this assessment. Yet the researcher believed that interesting insights into possible task problems, or needs for change, might be discovered from these students voices. According to Patton (1990), analyzing positive responses and negative responses can be considered extreme cases.

If students are asked specifics, they tend to give answers. Thus, question # 3 of the self-reflection provided a place for them to specifically write what they did not like in an attempt to understand their specific reasons. Students were asked, specifically, "What did you not like about the task?"

Eighteen of the "experienced" group and twelve of the "inexperienced" group when asked what they did not like responded "nothing" or they enjoyed "everything". Overall, this was the most frequent response to this question as thirty students or, 68%, wrote that there was "nothing" that they did not like.

The five (5) students, 1 in the "experienced" group and four (4) in the "inexperienced" group did not like the task because of the writing. Furthermore, even for some who enjoyed the Homer Booklet Assessment Task overall, when given the chance to respond to what part they do not enjoy, seven (7) students in the "inexperienced" group and one (1) in the "experienced" group did not like the writing. Thus, some students liked the task, in part, because of the writing, and some students did not like the task because of writing. It makes sense that those who enjoy writing would probably enjoy a task such as this, and those who don't enjoy writing would not.

Meaningfulness (Student Self-Reflection Question # 7--Worthiness of Knowing)

The final question of the student reflection was an attempt to understand children's perceptions about the importance, meaningfulness, or worthiness of the information itself. Gaining student's perceptions through the self-reflections might shed light about spending valuable physical education time engaging in time-

consuming alternative assessments. This section highlights student responses. The teacher's perception will be discussed in the feasibility section. The specific question on the student self-reflection was: "Do you think it is important to know about physical activity and health? Why? Why not?"

All students, regardless of instruction or not, completing this question (N = 44) believed that the information was important to know. Three themes emerged for both groups. Theme 1: The importance of being healthy for a lifetime. Theme 2: The importance of knowing how to stay healthy. Theme 3: Becoming a couch potato, living longer, or dying. More students in the "inexperienced" group wrote about living and dying, while more in the "experienced" group wrote about staying, being, and needing to be healthy for life.

Although, overall, J. B. did not like the Homer Booklet Assessment Task, he still felt the content of the information was important to know. In an early interview, this is what he said:

It's, um, very important because it will make a better, a difference on our life when we grow up. Because if we're not really fit when we're young because it might like make us get sick when we're older or be lazy when we're older because we, that's what we're used to doing--sitting inside watching TV.

Homer Feasibility Question

Results and Discussion

The fourth research question in this study was used to evaluate the assessment task according to feasibility. It was believed by the researcher that if an assessment is not feasible to administer in a regular physical education setting, it would not be valued by a teacher and, therefore, not ever used once developed. Consequently, this was an important question to the evaluation of the assessment for this study.

Patty found this assessment to be very feasible in a regular physical education classroom setting. She thought it was quick and easy and found the rubric clearly distinguished between those who knew the content and those who did not.

Patty: ... the booklet was easy--as far as to go back and look through it--to check to see if the four guidelines were on it. So, it did not take much time at all to go back and, you know, go through the material, for you know, the teacher part of it--just to check it off to see what they got out of it (referring to the instruction).

Patty also found the assessment to be authentic to the children since they had a real audience to write for. Patty said:

I think it made them stop and choose the words they wanted to put on there more careful than if they were telling me 'this is what I know'. They were thinking about teaching it to somebody else. I think they took the time to think.

Patty also thought the product could be elaborated on with classroom teachers to reinforce writing across the curriculum. Lee, one of the Content Experts also wrote: "Feasible--yes--but if you're looking for publishable quality, should be done in conjunction w/classroom teacher so kids can revise, use technology, etc."

The length of time needed to plan for and engage students in performance or product assessments is often much longer than a multiple choice or fill in the blank test. Therefore, time is one concern of alternative assessments. However, the Homer Booklet Assessment Task seems to be quick and easy. It does not take much time. In her interview, Patty thought of the Homer Booklet Assessment as an end of the period--10-minute assessment. Patty's own words help describe her perception of the time involved for developing an assessment task such as this one: "...something like the booklet wouldn't take--not that much time involved in putting that together--no more than me organizing on paper when I go to do my checklist."

Furthermore, Calvin, one of the Content Experts, uses these type assessments in his program to determine if students remember taught cues. The fact that he uses similar tasks reinforces the notion for the feasibility of this assessment in a regular physical education setting. Calvin said: "I do several writing assessments during the school year. These are the best at distinguishing those who cognitively understand cues for skill and movement and those who do not. They are more clear and objective."

In order for this assessment to be feasible for other teachers to use, Patty thought children would have to get used to writing during PE time. The first time they do anything takes longer. After they figure out the process and are used to doing it, things go smoothly, just like learning protocols. The teacher and Content Experts found this task to be feasible to administer in a regular physical education setting. They also felt the assessment could be enhanced further by working with a classroom teacher if "publishable" is a major focus for the final product.

Conclusion and Interpretation of Results

Students experiencing instruction scored high, while those not experiencing instruction scored low. Therefore, the assessment seems to discriminate between those taught and not taught. Content Experts agreed that the Homer Booklet Assessment Task had high content validity. This indicates that the assessment task is important to the subject matter, measures student learning, and has met the criteria of systematic construction for assessment tasks according to guidelines set forth by Herman, Aschbacher, and Winters (1992). A large percentage of the students found the task meaningful and even authentic, although there might be a difference between those receiving instruction and those not receiving instruction on this question. The Homer Booklet Assessment Task appears feasible for a regular physical education teacher to administer in a regular physical education

setting. Furthermore, the teacher and Content Experts thought the Homer Booklet Assessment Task might be even more feasible for producing publishable quality work when working with a classroom teacher.

Again, all findings should be evaluated within the context of the particular setting and should not be generalized to other programs and students outside of the context. Although an "expert panel" and "review team" collaborated together to make this study possible, their viewpoints may not represent other professional's interpretation of results.

CHAPTER 5

Results & Discussion

Video Tape Advertisement Assessment

The purpose of this study was to design three alternative assessment prototypes. This chapter describes the Video Tape Advertisement. The same four research questions guided the evaluation of the assessment task. They were: 1. Do students who have been taught the physical activity guidelines score differently on the alternative assessment than those who have not been taught the guidelines? 2. Do content experts agree that scores can be used to describe what students have learned (content validity)? 3. Do students find the assessment tasks worthwhile, enjoyable, and meaningful? 4. Is the alternative assessment prototype "feasible" for a teacher to use in a regular physical education setting?

This chapter contains sections including: 1. Video Tape Advertisement Assessment Task Description; 2. Score Results; 3. Content Validity; 4. Student Meaningfulness (enjoyment); 5. Feasibility; 6. Summary and Conclusions.

Video Tape Advertisement Assessment Task Description

For the purpose of this study, the Video Tape Advertisement Assessment Task is also considered a product assessment under the large umbrella term of performance assessments according to the National Education Association (McTighe & Ferrara, 1994). This task was a group assessment. With the help of the classroom teacher, the physical education teacher assigned children to groups of three or four students per group. Wherever possible, one student with strong verbal skills, and one student with strong writing skills was assigned to each group.

Students were asked to design and create a video tape advertisement for children their age that is both very interesting to them and full of correct information. They were told that most children their age do not know how to be healthy through physical activity for a lifetime. Therefore, a two to three-minute video advertisement could help children learn this information in an interesting way. A description of the task and analytic rubric was created to help youngsters with the process and expectations. (See Figure 7 Video Advertisement Task.)

The analytic rubric separated the task into two major categories; one for content and one for style. (See Figure 8 and Figure 9 for student and rater scoring rubric.) Two mean scores were then calculated from these two major categories. Each group mean score and the percentage of students scoring at all rubric score points are contained in the score results section. Two adults, the teacher and researcher, rated the Video Tape Advertisement Assessment Task.

Video Tape Advertisement Scores

Results and Discussion

The content to be included in this assessment was the same expected for the Homer Booklet. The four criteria are: 1) number of minutes per day, 2) intensity level, 3) number of days per week, and 4) select activities that are fun for you. The guidelines to be measured were pre-selected from the revised physical activity guidelines from the Surgeon General's Report (USDHHS, 1996).

During the initial pilot, students were asked their ideas for how to "show off" what they knew to other children. J. B. thought it would be active and fun to create a video tape to "show" other students their age how to be healthy through physical activity for a lifetime. Other students agreed. In addition, the Homer Booklet Task measured student understanding through the medium of writing.

Video Tape Advertisement Task

Background

Most children your age do not know how to be healthy through physical activity for a lifetime. A two to three minute video advertisement could help children learn this information in an interesting way.

Task

Your task is to design and create a video advertisement for children which is both very interesting to them and full of correct information.

Audience

The audience for your video advertisement is other fourth and fifth grade students at other schools.

Purpose

The purpose of your work is to use your knowledge of physical activity to entertain and teach children.

Procedure

1. Make a list of the guidelines or cues you will use.
2. Check the accuracy of your information. Be sure to explain the words.
3. Plan your introduction (beginning) and conclusion (ending).
4. Make a first draft of the graphics/pictures you will use.
5. Plan a sequence of the activity, graphics/pictures you will use.
6. Practice the order and sequence with your group many times.
7. Practice who will do which parts. Help each other do well.
8. Make revisions and complete the final draft of the graphics/pictures.
9. Make the video advertisement in front of the camera.

Figure 7. Video Tape Advertisement Task (Student Directions)

Performance Task Rubric
 Creating a Physical Activity
 Video Tape Advertisement for Children

Student Rubric Score:

- **** goes beyond what is asked with **quality and accuracy**
- *** clearly meets the criteria with **quality and accuracy**
- ** meets the criteria but lacks **quality**
- * partially meets the criteria--may lack **quality and accuracy**
- does not meet the criteria--**quality and accuracy is lacking**

Content Element	Group Score
1. The <u>4 cues</u> are included in the video advertisement [score by the number of cues included]	
2. Physical activity <u>vocabulary</u> is explained.	
Style Element	
3. <u>Pictures</u> and other <u>graphics</u> support the verbal statements.	
4. The introduction and conclusion catch the interest of the children.	
5. The story line is organized.	
6. The video advertisement is interesting and creative.	

Names of People in your group:

- 1.
- 2.
- 3.
- 4.

Figure 8. Video Tape Advertisement Rubric for Students

Video Tape Advertisement for Children Teacher and Rater Rubric

Teacher & Rater Rubric Score:

- 4** goes beyond the criteria with **quality and accuracy**
3 clearly meets the criteria with **quality and accuracy**
2 meets the criteria but lacks quality
1 partially meets the criteria--may lack **quality and accuracy**
0 does not meet the criteria--**quality and accuracy is lacking**

ASSESSMENT POINTS

Content Element	Group Score	Teacher Score	Rater Score	Mean Score
1. The <u>4 cues</u> are included in the video advertisement [score by the number of cues included]				
2. Physical activity <u>vocabulary</u> is explained.				
Mean Content score				
Style Element				
3. <u>Pictures</u> and other <u>graphics</u> support the verbal statements.				
4. The introduction and conclusion catch the interest of the children.				
5. The story line is organized.				
6. The video advertisement is interesting and creative.				
Mean Style Score				

Names of People in your group:

- 1.
- 2.
- 3.
- 4.

Figure 9. Video Tape Advertisement Rubric for the Teacher and Rater

The Video Tape Advertisement allowed students to be measured on the same content through two different mediums--verbal and kinesthetic. These two mediums are just two of six or eight processes for learning that Howard Gardner calls "multiple Intelligences" (1983) which he suggests are dispositions toward learning through different processes. Some children prefer kinesthetic means while others learn better visually or verbally, for example. Therefore, it was decided that a verbal (Video Tape Advertisement Assessment) and written (Homer Booklet Task) assessment would be designed, piloted, and evaluated for this study.

This section highlights rubric scores by group mean scores. This section helps answer Research Question # 1: Do students who have been taught the physical activity guidelines score differently on the alternative assessment prototype than those who have not been taught the guidelines? Figure 10 summarizes student scores.

Mean content scores for fourth grade were 3.1 for the "experienced" group and 0 for the "inexperienced" group. Clearly, students who were taught the content scored higher than those not taught. It is important to note that for this group assessment, scores of zero were possible in the content section as one content score simply asked the rater to count the correct number of guidelines included in the video. If no guidelines were included students scored a zero.

Students were also asked to explain any vocabulary they used for words other students might not know. For example, the word moderate was new to them prior to learning about it, so they were asked to think about how they would explain or show what that word meant. If students did not include, that or any other word needing an explanation, they received a score of zero.

Students were given full points for the second question of the Video Tape Advertisement content, if students showed with their bodies what moderate

intensity looked like--using an activity, or explained in words such moderate means "don't go too hard and don't go too easy" .

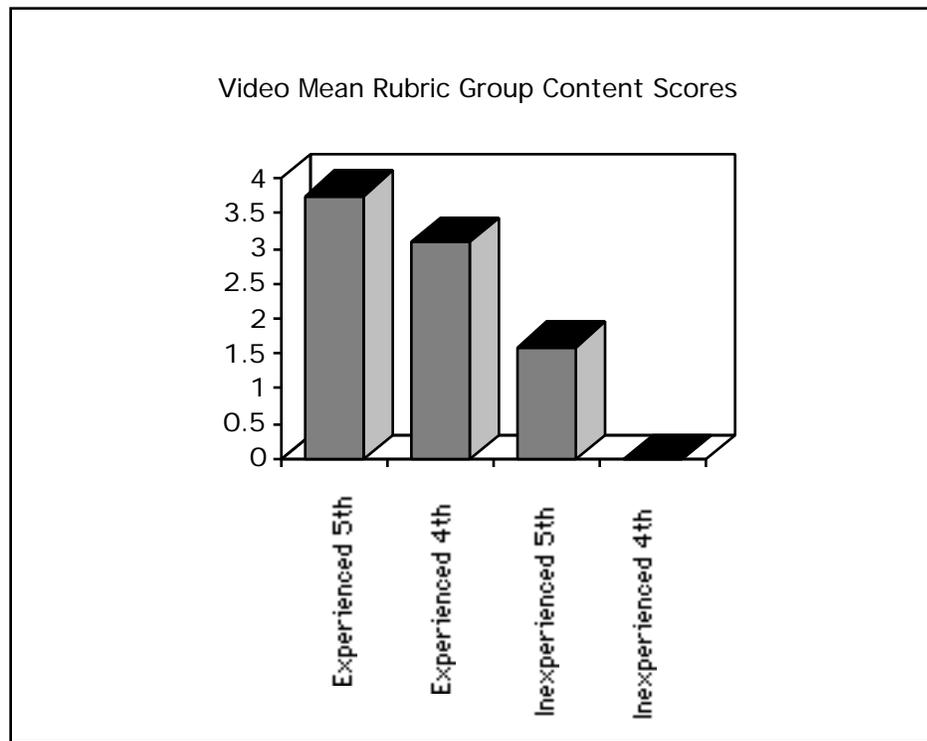


Figure 10. Mean content scores for the fifth grade "experienced" group was 3.7 and 1.6 for the "inexperienced" group.

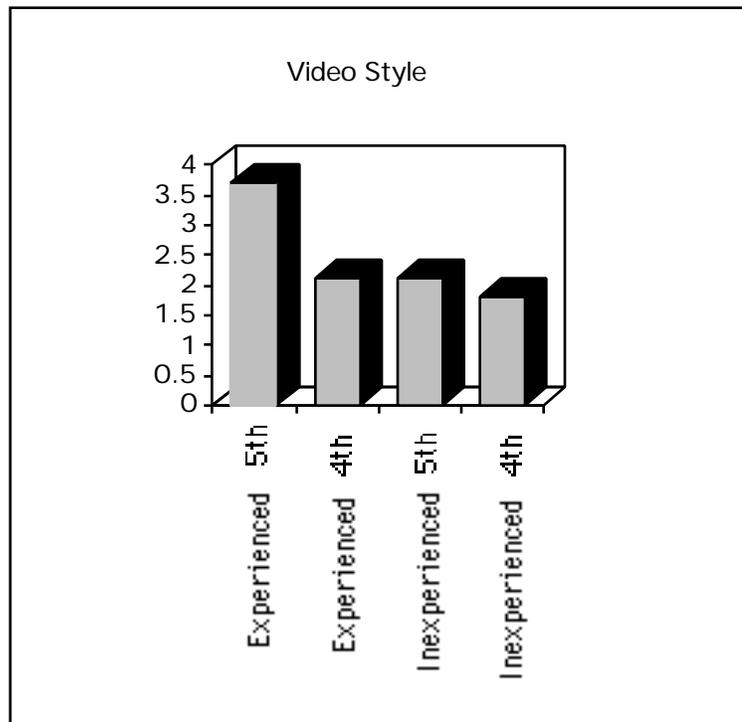


Figure 11. There was little, or no difference in style scores between fourth grade students who were taught and not taught. There was a difference with fifth grade.

It appears that the fifth grade "experienced" group scored better and was able to better complete the project according to the rubric and criteria. Field notes confirm these groups completed the project with more ease. It may be that fifth grade students are better able to work together, design, and follow a plan than fourth grade students. Once again this was a new task for all youngsters, and anytime students do new activities it takes more time and quality may be lacking until there are more student examples to share with them.

Although total mean scores for each intact "experienced" and "inexperienced" group provides an overall view of the scores, each group mean score might add

depth to the results. Table 4 specifically provides within group mean scores for the "experienced" groups. Table 5 specifically provides within group mean scores of the "inexperienced" groups. Mean scores were calculated by averaging both adult rater scores for content and style.

Table 4

Video Tape Advertisement Assessment: Within Group Mean Rubric Content Scores

"Experienced" Groups

Groups	5th Grade		4th Grade	
	Content	Style	Content	Style
Blue	4	4	4	2
Orange	4	3.25	3	2
Green	3.25	3.75	2	2.25
Red	-	-	3.5	1.75
Mean Score	3.75	3.7	3.1	2

*3 and 4 are considered high scores

*1 and 2 are considered low scores

Table 5

Video Tape Advertisement Assessment: Within Group Mean Rubric Content Scores

"Inexperienced" Groups

Groups	5th		4th	
	Content	Style	Content	Style
Blue	2	2.25	0	1.75
Orange	1.5	2	0	1.75
Green	1.5	2	0	1.75
Red	-	-	0	1.75
Purple	-	-	0	1.75
Mean Score	1.6	2.1	0	1.75

*3 and 4 are considered high scores

*1 and 2 are considered low scores

If scores of 3 and 4 are considered high scores versus 1 and 2 low scores, it is clear to see that 100% of "experienced" fifth grade students scored high and 100% of "inexperienced" fifth grade students scored low (see Table 4 and 5). The data indicates that for these fifth grade students, scores discriminated between those who received instruction and those who did not.

In contrast to fifth, only 75% (3 of 4 groups) of "experienced" fourth grade groups scored high while 25% (1 group) scored low. Yet, 100% of the "inexperienced" fourth grade students scored low. Taken together, 88% "experienced" fourth and fifth grade groups scored high. Furthermore, results indicate that 100% of students from both fourth and fifth grade scored low in the "inexperienced" group. So, no student in either fourth or fifth grade "inexperienced" groups scored high. Scores, however, appear to discriminate better for fifth grade students than for fourth grade students.

It is believed that the same three individual students with prior knowledge were in those "inexperienced" groups where students scored 2's. It is also possible that fifth grade students learned the moderate or medium intensity level guideline from doing the Perceived Intensity Level Assessment. For fifth grade classes, the Perceived Intensity Level Assessment Task was completed before the Video Tape Advertisement Assessment Task which came last. Fourth grade, however, completed the Video Tape Advertisement Assessment Task first and it appears some groups were not yet proficient with the content.

The last task for fourth grade students was the Homer Booklet and mean scores indicated that they scored better than fifth grade on the booklet. (See Chapter 4). Fifth grade students completed the Homer Booklet Assessment Task first. This might indicate that by the time the last assessment is administered, students are more confident and efficient at applying the information they learned through instruction and possibly the assessments.

Video Tape Validity Question/Content Experts

Results and Discussion

An important validity question was just addressed in the previous section. Do scores discriminate between students who have been taught and students who have not been taught? If there are no differences, something is wrong with the assessment. The following section focuses on Research Question #2. Do content experts agree scores can be used to make judgments about what students learned? This is also an important content validity question.

This section highlights important reliability and particularly validity questions. Inter-rater reliability scores were reported to the content experts. Inter-rater reliability for the Video Tape Advertisement Assessment Task was 94% for two raters at the perfect match level. Inter-rater percent reliability was determined by the number of matches divided by the number of possible matches for the analytic scoring rubric. The two adults matched 90 out of 96 possible score points at a perfect match. It is important to note that raters agreed the Video Tape Advertisement Task was difficult to score, and they had to watch the video tapes several times to score all the criteria.

In an attempt to determine if, or how, scores can be used to make judgments about student learning, the same three content experts evaluated the Video Tape Advertisement Assessment Task. All student video assessments were available for content experts to watch. Furthermore, all teacher and student task directions and scoring rubrics, including all changes from beginning to end, were also available for content experts to examine. Again, the three content experts were asked to rate the assessment based on the same thirteen questions, using the same rater scoring rubric. See Figure 12 for the content expert scoring rubric.

The three Content Experts gave scores of 2's on all but five questions that were scored as 1's. Calvin scored question # 4 a 2 for content, but 0 for style. Using

the 0 score, the average overall rubric score for all the questions by the three outside Content Experts was 1.8 out of a possible score of 2. Therefore, the Content Experts agreed that the Video Tape Advertisement Assessment design satisfactorily met the criteria for content validity. Based on one content expert, however, it appears that there are some style concerns that need to be addressed and possible modifications needed before the task and rubric are used with other groups.

Written comments from the Content Experts help deepen the understanding of their judgments. The following comments are their insights about the Video Tape Advertisement Assessment Task Scoring Rubric. Lee wrote: "I think it made the rubric very clear to state that the physical activity vocabulary be explained. Really a good idea to separate content element and style element."

Calvin noted that "scoring the content seems more objective and clear. Style seems quite subjective." This comment was written in response to the way he scored question number four with style and content separately. His insight is perceptive. Professional judgments must be made in scoring these types of tasks and, therefore, rubrics need to be developed to make scoring judgments as systematic as possible. Rubrics should be made public so students and teachers know the criteria for distinguished work versus poor work. Over time, with many student samples, the teacher should notice an improvement in the quality of student work.

Joe wrote: "the form of assessment would not be suitable for all children given different level to communicate verbally." This is an important insight and indicates his agreement with Gardner (1983) that children have different learning dispositions, processes, or "multiple intelligences." It is important to note, that for this very reason, the physical education teacher conferred with the classroom teachers to make sure, wherever possible, one

Content Validity Rubric (Video Tape Advertisement)

<p>Rubric for rating:</p> <p>2 Yes, clearly meets criteria without hesitation or reservation.</p> <p>1 Yes--but depth or quality is lacking, and revisions may be necessary.</p> <p>0 Cannot answer yes to the question, therefore does not meet the criteria to the question.</p>
--

Validity Questions: Can the scores be used to describe what students have learned?	Rubric Score
1. Does the content of this assessment reflect important content in physical education curriculum?	
2. Is there a good match between the skills and knowledge intended for this assessment and those emphasized in this PE program?	
3. Is there a clear definition of the assessment goal or purpose so that judgments can be made about the match between knowledge and understanding intended for this assessment?	
4. Does the assessment task measure student's understanding of physical activity?	
5. Does the assessment task matches the level of thinking necessary to assess the kind of information sought?	
6. Does the assessment task include scoring criteria?	
7. Does the criteria include standards for judging the student performance?	
8. Does the rubric or scoring criteria appear to discriminate between those who understand and those who need further assistance?	
9. Is the task developmentally appropriate for fourth and fifth grade students?	
10. Does the assessment task reflect processes and outcomes suitable for the fourth and fifth grade students?	
11. Did students have sufficient opportunity to learn what is included in the assessment?	
12. Does the assessment task require students to apply their understanding of physical activity in a new situation?	
13. Does the scoring rubric accurately describe student's understanding of the guidelines for how to be healthy through physical activity for a lifetime?	
Average Rubric Score for the validity of this assessment	

Is this a feasible assessment for a regular physical education class? Why, why not?

Figure 12. Content Validity Rubric (Video Tape Advertisement)

student with strong verbal skills was included in each group for this group Video Tape Advertisement Assessment Task.

In order to further elaborate on content validity and task concerns, Patty's insights and comments are included next. These insights are from her interview.

Researcher: Okay does anything need to be changed on the directions or the task or is there a way to make the rubric more clear to make it easier to score? Or is it just something that if a teacher chose to do that and the kids were used to it...maybe a performing arts school where they do acting.

Patty: Well as far as the rubric--I mean we played around with it several times tried to get it to where it would work--to where it was clear. I think that the last one we used, you know, I think it fell into place. I felt like it was pretty clear. (See the difference between Appendix G Video Tape Rubric and Figure 9 Video Tape Rubric).

Researcher: Better

Patty: Some of it we still played around with the wording on it accurate information but quality was lacking and I think there needs to be a small stipulation in there that the information was there, but the quality could be improved. And that would be the only stipulation on that one that I would look for because you can satisfactorily do something, but you can also do something on a good or an excellent level. You still got it, but they could improve the quality so that would be better. (See Figure 9 for revision).

Patty's comments indicate that the rubric needs to be modified with one more level of quality before it could be used again. In our team's opinion this would help to better discriminate between the quality of the work. (See Figure 9).

The first question for content in the analytic rubric asked the rater to score the correct number of guidelines included in the video. The second question asked the rater to score the quality of student understanding by student's ability to explain the vocabulary or guidelines used. This was meant to make deeper judgments about student's understanding. Students who were taught the content scored higher than students who were not taught. Therefore, that part of the analytic rubric is probably not a concern. However, the style section may be a concern.

Once again, it is important to point out to the reader that extreme caution should be used with these results. These scores are from small groups of youngsters from one physical education setting, one teacher's program, and four teacher content experts. These assessments and rubrics would need to be further piloted and evaluated in many diverse settings with lots of youngsters before true judgments could be made about them.

Student Meaningfulness (Enjoyment) Question

One of the criteria for tasks leading to sound assessments is the notion of meaningfulness. The Center for Research on Evaluation, Standards, and Student Testing (CRESST) sets general criteria for evaluating assessment tasks. One of the criteria is meaningfulness. Herman, Aschbacher, and Winters (1992) include the CRESST criteria in their book that was used in the design of the assessment tasks. They ask the question: "Will the tasks be meaningful and engaging to students so that they will be motivated to show their capabilities? Do the tasks involve real problems, situations, and audiences?" (p. 42). In this study, the third research question asked that question in order to measure students' perceived enjoyment, meaningfulness, authenticity of the assessment task.

Again, when students were interviewed it became clear that the word meaningfulness did not mean much to them. Enjoyment, however, was a word children understood and could relate to therefore enjoyment was used as an indicator of meaningfulness to the youngsters. Furthermore, the last student reflection question asked students if they thought the information they were learning was important to know. The fact that students perceived the information as important to know was also interpreted by the researcher as one indicator of meaningfulness to the students.

Results and Discussion

Table 6 displays simple number counts and percentages for student responses to the first student-reflection question completed after the Video Tape Advertisement Assessment Task. This question was: "Overall, did you find this assessment to be enjoyable? Why? Why not?" Students answered yes or no, with one okay response.

Forty-six of forty-seven, or 98% of the students, enjoyed this task. One student did not like it, and one student indicated it was okay. The okay response was included in the yes percentage counts. Thus, students, whether taught or not, found this task very meaningful. Most found it fun and they enjoyed being on camera. Student self-reflections, interview data, and the final survey results confirm that students found this task to be enjoyable. The researcher translates enjoyable into meaningful for the student.

Table 6

Video Tape Advertisement Assessment Overall Task Enjoyment

Groups	n	<u>Yes</u>	<u>No</u>
5th Experienced	11	11 (100%)	0
5th Inexperienced	12	12 (100%)	0
4th Experienced	15	15 (100%)	0
4th Inexperienced	9	8 (89%)	1
Totals	47	46 (98)	1

Note: The 1 okay response is included in the yes category
 Yes = # of students who enjoyed the assessment task
 No = # of students who did not enjoy the assessment task

Meaningfulness (Student Self-Reflection Question # 1 & 2)

Four themes emerged for students who found the assessment enjoyable. The self-reflection question specifically asked: "Overall, did you find this assessment to be enjoyable? Why? Why not?" Three data sources (self-reflections, interviews, "Your Favorite Task") indicated students really enjoyed the Video Tape Advertisement Assessment Task. Themes from student self-reflections included: Theme 1: Students enjoyed making the video, being on camera, or TV. Theme 2: Students found it fun or different. Theme 3: Students enjoyed working with or helping others. Theme 4: Students enjoyed being active while making the video.

Quotes are typed in exactly as found in the student reflections including spelling and punctuation. Interview data is also used to support the themes. Students found the assessment enjoyable because...

Theme 1: Students enjoyed making the video, being on camera, or TV.

"Yes. because It might be on tv" [SR: Leon VC4]

"Yes because we got to make a cumershul" [SR: Hidi VH5]

"Yes because it was fun We got to be on camera" [SR: Jes: VS4]

"I think it's fun to be on video tape and see yourself." [I: Rodney VS5]

Theme 2: Students found it fun or different.

"Yes, because fun an helpful" [SR: Em VS4]

"Yes, because it's some thing different that we haven't ever done it before."

[SR: Sean VS4]

"Yes because it was different and fun." [SR: Julie VH5]

When asked which task he enjoyed the most, Harold said:

The video, because our group, what we did is, um we appointed like different people to do different tasks. And see, I was the, um, stage manager and everything, and I got all the equipment set up. I did a bunch of other things, and I also got to be in the video. That was very exciting, and good, and, um, I thought the um...We did a lot of practice on the video so it helped us out really good. [I: V Harold VH5]

Theme 3: Students enjoyed working with or helping others.

"Yes! I enjoyed it because we got to do something to help other people." [SR: Mandy VC4]

"Yes because we got to show people how to be healthy" [SR: Anna VH5]

"Yes because I liked it because I worked with people" [SR: Matt VH5]

"I enjoyed the video the best because I got to work with my friends and we got to learn more and just do stuff and show the younger people what you should do to stay healthy for a lifetime." [I: Kattie VC4]

Theme 4: Students enjoyed being active while making the video.

"Yes because we got to do stuff as we did it." [SR: AJ VS5]

"Because it wasn't just recording it was physical activities." [SR: Kev VC4]

"Yes it had lots of stuff like gymnastics and basketball." [SR: Alan VS4]

In an interview about the Video Tape Advertisement Assessment Task, Lee and Kyle were asked why they enjoyed it. The following was their response:

"I liked doing it because you would use all the knowledge you learned and understood in your own words and then instead of writing it you got to act (said with emphasis) it out." [I: Lee VS5]

"The video because we practiced the activities and got to play the stuff." [I: Kyle VH5]

From student's spoken or written words it is clear to see they were very enthusiastic about the Video Tape Advertisement Assessment Task. Making the advertisement, using their own words, and helping others all indicate authenticity. Whether perceived or real, it seemed to heighten their enthusiasm toward completing the task.

Meaningfulness (Student Self-Reflection Question # 3)

In this study, students evaluated the assessment tasks according to their overall enjoyment. Because of such positive results, the researcher also decided to

analyze negative and okay responses. Only one student responded that he did not like this assessment task. One girl thought it was just okay. Yet the researcher believed that interesting insights into possible task problems, or needs for change, might be discovered from these student's voices. According to Patton (1990), analyzing positive responses and negative responses can be considered extreme cases.

The student who responded no, and the student who responded okay, both indicated they did not like being in front of the camera, or in front of their peers. In fact, the student who found it okay experienced a contextual barrier. The group had designed and practiced for two days, and on the fourth day when it was time to tape, their group's speaker was absent. None of the other students in the group wanted to be the speaker. Jennifer ultimately compromised with her group and agreed to be the speaker, but it made her uncomfortable. Despite these barriers, this group scored a 4 on content and a 3 on style. This concern for speaking in front of others is very real to some people, children included. So, it makes sense that there will always be some students who are uncomfortable with one medium or another for demonstrating what they know and can do. That is why it is important for the teacher to vary the assessment tasks in order to obtain the best results over time for all students.

Video Tape Feasibility Question

The fourth research question in this study--"Is the alternative assessment prototype 'feasible' for a teacher to use in a regular physical education setting?"--was used to evaluate the assessment task according to feasibility. It was believed by the researcher that if an assessment is not feasible to administer in a regular physical education setting, it would not be valued by a teacher and, consequently,

not be used once developed. For this study, therefore, feasibility was an important evaluation question.

Results and Discussion

The teacher found this task to be the least feasible of the three. It involves technology such as cameras and microphones that not all schools have.

Furthermore, it is very time-consuming to measure information that could be more efficiently obtained by other means such as the Homer Booklet Task. Patty discusses her point of view in the teacher interview.

Patty: This one took probably the most time to do which I think it was a neat way to get across the information...I think the kids had fun producing it...when you got them up in front of the camera they froze (laughs)...they were a lot more flamboyant and stuff about it in their practice. I think it was kind of fun, but I don't think it would be one that I would choose, um, just the amount of time that was involved.

Researcher: How many days do you think it would take to do it well? We took four.

Patty: Well most...several of the kids had it ready in 3 days, but the video taping was the part that slowed it down...having to use that fourth day to finish video taping. But the kids pretty much had the information down...now if this would be something that they were accustomed to doing, they would probably be better at organizing their thoughts and getting together a little quicker and then with being a video taper the more I practiced it, it might be something that would just take a day or two. But you know, but the first time, I'm just like, well, it was a lot of time...I'm not sure whether I'd want to do this one or not. The booklet was short and quick, gave me the information that I was looking for whether they had it or not...It was pretty cut and dry...The video tape was hard to score. It was...the performances were so short they were hard to critique. The ones that had the information...it was real cut and dry...the ones that were a little shaky on it, you , you want the information to be there so you try to search for it hard, which makes it really hard. It was just a little bit more difficult. I think as far as if we were doing this special presentation thing specifically for the purpose to send it someplace else, but to just use it as an assessment for me I don't think that to go to all the trouble and all that is a lot of work involved getting the video camera making sure it works and the microphones which, we don't have one for starters, in an elementary school.

Furthermore, if students were used to this type of assessment or activity it might take less time. Patty said:

...as I said, if you practiced that type of thing and video taped them often, and let them look at themselves like video tape them throwing and say okay give me the throwing cues how many of the cues do you see yourself doing? But I think once they got used to themselves being in front of the camera, which I think all schools would love to have that kind of equipment to do those types of things, I think video taping and making a commercial they would be a little bit more comfortable being up in front of a camera and be a ham. And they would then show their personalities, because some of them really have nice little personalities, but we didn't see it. While they were practicing we did. But when the camera got turned on it was like this is Dan Rather in front of the TV. It was hilarious but um...

In order to make the Video Tape Advertisement Assessment Task a more feasible and efficient assessment task, the researcher would change a couple of procedures, such as creating specific roles for the students, and utilizing an outside person do the video taping. One of the student groups came up with the notion of creating specific roles for every person in their group. Unfortunately, they were in the fifth grade group which taped last, so we didn't learn from them in time to implement the change. This task, however, could be modified to include their idea. For example, there could be two equipment managers--one for PE equipment, and one for materials such as markers, folders, etc.; a speaker; a director and "keep us on task person"; and a writer for the graphics. This would give individuals within the groups thereby allowing them to be more successful with the style of the presentation. Implementing this change would provide students with more direction about what they needed to do.

Secondly, by utilizing someone outside of class to tape the groups the teacher would not lose so much instruction time. This was also recommended by two of the content experts. The video tape operator could be a teacher's aide, a student teacher from a classroom, a parent, older student, or other volunteer. This way the teacher could take only two days of instruction time for the Video Tape

Advertisement Assessment Task. During those two days, students would design and practice their video advertisement presentations. During the next two days students could be pulled from the physical education class, by groups, to be taped. Thus, the physical education teacher to continue teaching during those two valuable instruction days.

An important question for the Video Tape Assessment Task is: If the content is the same, and the Homer Booklet takes ten minutes of the teacher's instructional time, whereas the video takes at least four teaching days--two for design and practice and two for taping--is doing the video for this type of knowledge worthy of the time it takes to administer? Each teacher would have to make that decision. It would seem, however, that other content--such as students performing an overhand throw-- might be more worthy of the time needed for a video assessment rather than content that can be captured in less time through other means such as a booklet.

Feasibility seems to be a primary concern for this Video Tape Advertisement Assessment. As previously mentioned, if the content were different, it might be worth the time involved. Certainly, there are pros and cons with any assessment, and it is always important to match the assessment with the purpose, knowledge, and level of understanding desired.

Conclusion and Interpretation of Results

In this Video Tape Advertisement Assessment, students who were taught the content scored higher on content than those who were not taught. Style scores were not as discriminating between those taught and not taught. Students who were taught scored higher on mean style and content scores taken together. Therefore, scores do discriminate between those taught and those not taught. However, fifth grade scores may have discriminated better than fourth. There

could be a developmental difference between fourth and fifth grade, or the fact that fourth grade completed the Video Tape Advertisement Assessment Task first, while fifth grade completed it last, could have been the difference. Fifth grade had more opportunity time to remember the guidelines for this assessment.

Content Experts agreed that this assessment had high content validity. There may be a problem, however, with style as it is more subjective. It would be important to train raters for consistent scoring, and it is very important to refine the rubric as student results become available.

Students found this assessment task the most enjoyable. Only one student did not like it, and only one found it okay. Those two students did not like speaking in front of others, or the camera. The changes recommended by the researcher, found in Chapter 7, would appear to eliminate this variable for shy students, or students not comfortable with speaking before an audience.

The teacher found this assessment to be the least feasible because of the equipment needed and time it took to obtain information that could be obtained through other means. Content experts found it to be feasible to do in regular physical education classes, if some changes were made--primarily finding a responsible person to do the actual taping so they could continue to teach.

Ultimately, teachers must make tough decisions about both content and assessment in their programs. Results from this Video Tape Advertisement Assessment Task indicate that a teacher could show evidence of student learning, but, the task might be more involved than necessary for this particular content. Yet, if authenticity and engaging student's interest is important or necessary, this or another Video Tape Advertisement Assessment Task, might be a great choice.

The reader is again reminded that results from this study should not be generalized to other settings. These tasks need to be evaluated with numerous children, in numerous settings, before any final judgments can be made.

CHAPTER 6

Results & Discussion

Perceived Intensity Level Assessment Task

The purpose of this study was to design three alternative assessment prototypes. This chapter describes a Perceived Intensity Level Assessment Task that was designed to measure student perceived intensity levels. The same four research questions guided the evaluation of the assessment task. They were: 1. Do students who have been taught the physical activity guidelines score differently on the alternative assessment prototype than those who have not been taught the guidelines? 2. Do content experts agree that scores can be used to describe what students have learned (content validity)? 3. Do students find the assessment tasks worthwhile, enjoyable, and meaningful? 4. Is the alternative assessment prototype "feasible" for a teacher to use in a regular physical education setting?

This chapter contains the Perceived Intensity Level (PIL) Assessment Task Description, Score Results, Content Validity, Student Meaningfulness (enjoyment), Feasibility, and Conclusion & Interpretation of Results.

Perceived Intensity Level (PIL) Assessment Task Description

The PIL task was different from the other two assessments. This was a performance assessment that required students to be active at stations, for example, jump rope for one minute. Next students judged, then coded their perceived intensity level on their "Circle Task Coding Sheet" before rotating to the next station. Specifically, students coded themselves as having worked easy, moderate, or hard--(E, M, or H) at each station. Students continuously rotated to four different aerobic type stations. They repeated the four stations as time permitted. See Perceived Intensity Coding Instrument, Figure 14.

The second day of the assessment, fourth grade students participated in exactly the same stations, but coded both themselves and a partner. This was done to determine if, or how well, students could accurately code the perceived intensity level of someone else their age. Theoretically, if a student and teacher match well on coding of that student or a third student, accurate performance measurements would be possible in a regular time setting to document student learning. Consequently, for feasibility, an outside coder would not be necessary in order for teachers to successfully document reasonably accurate, for example, intensity levels of their students.

The teacher and researcher coded the students at the jump rope station. Typically three or four students were coded at a time on the one minute interval. Sixty-two students participated in this assessment task. Student understanding of perceived intensity level was measured by the number of student coding responses matching an adult coder at the same station. This assessment did not measure understanding of the other three guidelines. See Figure 13 for PIL Final Teacher Designed Task directions.

Many refinements, to both the task and coding instruments, were made during the pilot phase of this task. For example, designing the task score sheet so students could code as easily as possible was time-consuming. Many changes were made before the final score sheet using circles was found to be the most efficient and easy to understand (See Figure 14 for final circle task coding instrument for PIL for the student).

The coding system also changed. Initially, we tried a coding system from the SAPAC (Self-Administered Physical Activity Checklist) instrument found to be fairly reliable with children at this age (McKenzie et al., 1996). Self-reported physical activity has been found to correlate moderately with objective activity measures and should be considered to be rough approximations of true physical

activity (Sallis et al., 1996). The instrument had been found to be generally accurate as compared with objective measures such as heart rate monitors. Using this instrument to estimate their intensity level, students code themselves as having worked at a moderate intensity level: (M) most of the time, (S) some of the time, or (N) none of the time.

In follow-up interviews, the researcher discovered that this particular coding method was difficult for these children. While some students found coding okay, or easy, many thought it would be easier to code themselves (E) for easy, (M) for moderate or medium, or (H) for hard intensity level. Students were familiar with these terms and it, therefore, appeared to be a better choice for them. We were simply trying to measure rough approximations of student understanding of their personal intensity level, so this coding procedure was acceptable for this study. (See Appendix H, Figure 1 and 2 for coding changes).

The PIL Assessment Task also changed from fourth to fifth grade intact classes. First, fifth grade created their own stations, but we found that even in the "experienced" group some students did not create appropriate tasks. For example, a couple of groups simply created throwing stations which were not active enough for participants to move at moderate or hard intensity levels. Continuous movement with large body parts needed to be possible at the created station. Space was also a concern. There was not enough room in the multipurpose room for all the stations that students had created. Time was also a factor. Designing anything always takes more time.

Although the PIL Task, where students designed their own stations, was an activity students enjoyed, it did not match the ultimate purpose. Furthermore, some groups kept changing their activities and wanting different equipment. Ultimately, we wanted to know if students at this age could code themselves accurately, so creating the station did not really match the intended purpose.

Therefore, the teacher and researcher decided to create four stations all students could rotate to and students would simply code. The teacher designed task better met the purpose of the assessment (see Figure 15).

Fifth grade completed the Perceived Intensity Level task two ways; student designed and teacher designed. Fifth grade, however, did not do partner coding. Fourth grade only used the final coding instrument and teacher designed Perceived Intensity Level Task activity. Fourth grade students coded themselves the first day, and themselves and a partner the second day of the assessment. The teacher and researcher discovered their coding was more consistent at the jump rope station. Therefore, the jump rope station was the station used to watch and record students as they worked. Children's coding sheets were later compared with the adult raters to determine the percentage of matches for the group.

The teacher and researcher coded in real time, as opposed to watching on video-tape, while observing students. Percentage agreement was used to determine inter-rater reliability at the perfect match level. Within one score point, raters agreed 100% of the time. But, there was only a three point scale, so perfect match level was used. Perfect match level means the raters agreed perfectly on the rating. For example, they both scored a particular child as E for easy.

Students coded themselves inside the circle. The fifth grade groups that coded a partner coded themselves in the circle and their partner on the horizontal line next to the circle. They wrote E for easy, M for medium, and H for hard intensity level. Students wrote the name of the station above the first circle so that the researcher knew where students started and could confirm that student and teacher coding was checked at the jump rope station.

For the self-coding Perceived Intensity Level Assessment, inter-rater reliability (perfect match) was found to be at acceptable levels for all groups, but increased in accuracy with each group. According to CRESST rater agreement

guidelines (Herman, Aschbacher & Winters, 1992), exact agreement percentages can drop to 75 to 80% agreement. Specifically, inter-rater percentage consistency for the teacher designed Perceived Intensity Level self-code was 75% with the first group, followed by 77%, 81% and 94% with the final group. Because inter-rater reliability was found to be at the low end, especially with fifth grade, caution should be used when interpreting the results.

The question becomes: Although acceptable, if the raters did not agree, how could we know if students were accurate in matching the raters. Therefore, in the results section percentages were calculated based on the student matching one or both raters, but not necessarily both raters. In other words, if the student matched one rater that was counted as an accurate match for perceived intensity level.

Student Station Task (Final Teacher Designed)
Perceived Intensity Level Assessment Task

Background

For this task we want to help you learn what moderate intensity level feels like, so you can maintain that level for at least 10 minutes at a time. A good way to check this is the talk-sing test. If you can still talk to someone but can not sing you are working at the moderate level.

Task

We will show you the station activities to do. We have picked activities that will keep your arms, legs, and large muscles moving continuously, the whole time. You will rate yourself and your partner's intensity level for each one minute interval.

Audience

You will be an audience for yourself, your partner, the teacher, and one other adult rater.

Purpose

The purpose of your work is to see if you can rate your own intensity level correctly. It will also measure whether you can rate your partner's intensity level correctly. The teacher and one other person will also rate your intensity level. We will check to see if the adult rating matches your rating. Therefore, it is important to try to judge yourself and your partner fairly and correctly.

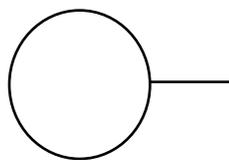
Procedure:

1. We will demonstrate each activity level so you can see the difference.
2. You will listen to the signal so you know when to be active and when to code.
3. We will demonstrate each station for you.
4. You will practice coding students we select to work at different levels.
5. You will get to practice being active and coding yourself and your partner.
6. We will begin and you will rotate to each station after you do your station activity 1 time.
7. Above the circle, code the first letter of the station you are at. For example, if you are starting at the jump rope station, write a **J** above circle number 1. We will show you how to do this.

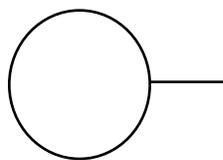
Figure 13. Perceived Intensity Level Assessment Task (Teacher Designed--Final)

Student Circle Coding Sheet

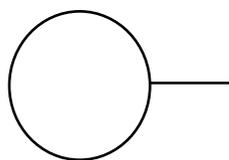
Your Name: _____ Partner's Name: _____



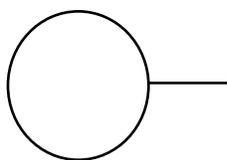
Practice



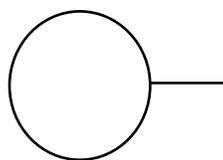
Practice



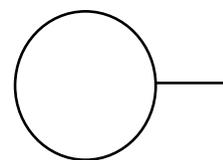
1



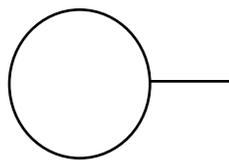
2



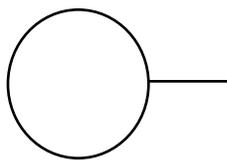
3



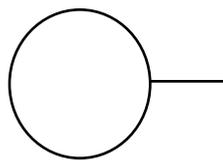
4



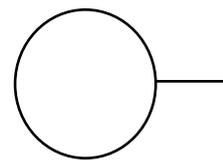
5



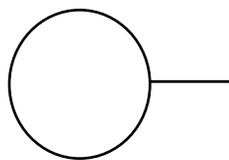
6



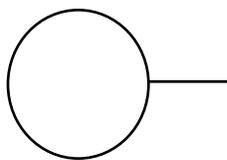
7



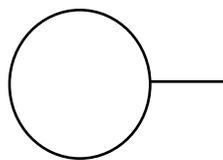
8



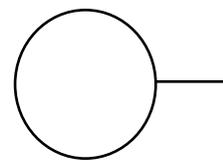
9



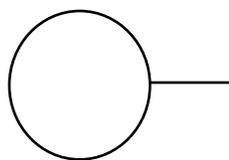
10



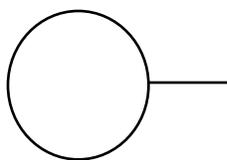
11



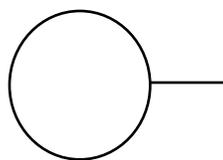
12



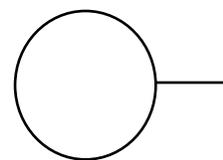
13



14



15



16

Figure 14. Student Circle Coding Sheet--PIL Assessment Task

The following codes originally fit on the Student Circle Coding page as a footnote. In this document, however, they do not fit on Figure 14.

Code yourself inside the circle:

Code your partner on the line next to the circle:

H = Hard or vigorous level most of the time

M = Moderate level most of the time

E = Easy level most of the time

Teacher Designed Stations

1. Basketball Station

Dribble the basketball around the cones. Keep going.

2. Skating Station

Turn the carpet squares upside down so they will slide on the floor.

Skate up to the cones and go around them. Keep going.

3. Hula Hoop & Catch Station

Both of you must stand inside the hoop. Pass the ball to your partner. Once your partner catches the ball both of you run switch hoops. When you are both inside the hoop your partner tosses the foam ball to you. You both run and switch again. Keep going. You always pass the ball from the same blue hoop--never the red.

4. Jump Rope Station

Jump rope in your self space at this station. Use a rebound bounce. If you can not turn and jump, place the rope on the floor and jump back and forth over it.

5. Walk The Dog Station (if we need)

Walk around the perimeter of the gym. Try to sing. If you can sing you are not working hard enough. If you can talk but not sing you are probably working at the right level.

Figure 15. Teacher Designed Station Tasks

Perceived Intensity Level Task Scores

Results and Discussion

In this study, it appears that both fifth and fourth grade students, regardless of whether they had instruction or not, were able to estimate their own perceived intensity level (H, M, or E) with one or both adult raters at the time they were coded. See Table 7 for results.

It appears that, regardless of instruction, youngsters are able to make fairly accurate judgments about their intensity level, as compared to a professional judgment by a trained adult. The percentage scores may be deceiving, however. Note fifth grade for example. In both classes only one student did not agree with the rater, however, the percentage is different because of the different number of students in each group.

It also appears there may be a difference in scores between fourth and fifth grade student's matching themselves with an adult rater. This difference could be due to a developmental difference between fourth and fifth grade students in understanding intensity level.

Partner Check Results and Discussion

Only fourth grade completed the second assessment. This assessment asked students to score themselves and a partner during the coding. When combining all possible agreement numbers, (52/63), inter-rater agreement for the self and partner task was 83% at the perfect match level. Agreement was 91% (32/35) at the perfect match level for the "experienced" group of fourth grade students and only 63% (20/28) for the "inexperienced" group. Thus, adult raters did not match each other well on their coding of students. Inter-rater agreement is below acceptable levels for the "inexperienced" group.

Reasons for this might be that one student in the "inexperienced" group got tripped by a ball rolling into her area. She hurt her ankle during the coding. At a

later time, a second student experienced a physical problem. Both times all activity and coding was stopped. It is believed that the teacher and researcher were distracted with their duties because contextual variables that took precedent when those two incidents occurred. They may not have been as careful in their observations. This might be the reason inter-rater reliability was so low for this group. See Table 8 for fourth grade self and partner matches with adult rater.

Table 7

Student Self-Match with Adult Rater

Percent & Number of Matches

<u>Grade Level</u>	<u>Experienced %</u>	<u>Number</u>	<u>Inexperienced %</u>	<u>Number</u>
Fifth	88%	15/16	92%	12/13
Fourth	71%	12/17	63%	10/16

Table 8

Fourth Grade Self and Partner-Match with Adult Rater

Percent and Number of Matches

	<u>Experienced %</u>	<u>Number</u>	<u>Inexperienced %</u>	<u>Number</u>
Partner code	73%	11/15	75%	12/16
Self code	100%	15/15	81%	13/16

From Table 8 it is clear that fourth grade students, whether taught or not, appeared to accurately code themselves. They appeared to be less accurate in coding their partners as compared with adult raters. There may be a difference in the self code for these fourth grade students, however, both groups tended to score high. During this task, it also appears that there may be a difference between

coding self and coding a partner as compared to an adult. Results are difficult to interpret because of low adult rater agreement for the "inexperienced" group.

Adults raters did not match nearly as many times with the "inexperienced" group. Therefore, more students in the "inexperienced" group had a chance of matching one of the raters as two different coding levels were often found with this group. This may indicate that there is a difference with the self code since 100% of the students in the "experienced" group matched with one or both adult raters for their self-check. Since adult raters matched 32 out of 35 times with the "experienced" group, matching was probably more difficult for the "experienced" group.

Perceived Intensity Level Validity Question/Content Experts

Results and Discussion

An important validity question was just addressed in the previous section. Do scores discriminate between students who have been taught and students who have not been taught? The results indicated that most students tended to score well. Thus, it appears there is no difference in this assessment task between those students who were taught and those who were not taught. However, it is still important to discuss what Content Experts thought. Therefore, the following section focuses on Research Question 2. "Do Content Experts agree scores can be used to make judgments about what students learned?" This is also an important content validity question. This section also includes the teacher's perspective.

Commenting of the appropriateness of the assessment and the design matching the purpose Patty said: "Getting across the concept of intensity, the stations were a perfect way to do it."

Researcher: Could you think of that assessment as a learning experience, actually? Like, one of the theories of alternative assessment is that teaching

and learning are more closely connected, that the assessment will actually help a child learn.

Teacher: I think that.

Researcher: that the teacher and student are learning together?

Teacher: Yeah, I think the assessment did help several of them. By the time they were finished, course you and I both heard comments from the kids, about...

Researcher: Quoting a youngster: 'I never sweated so much in my entire life.'

Teacher: Right, 'I started out too hard.' Didn't do that moderate level and they saw the difference in just a ten minute stint is all it was. So that directly hit home, showed them what their body was doing in a hurry.

Content experts also commented that they felt this Perceived Intensity Level Assessment Task might be a better instructional activity, versus an actual performance assessment. Since scores did not discriminate well, it might be better used as an instructional tool to help youngsters learn about pacing and intensity level. By describing the task, how it was taught, and what students did, Content Experts came to understand that the youngsters seemed to learn from 'doing' this assessment. Whether in the instructed group or not, children seemed to come away from the lesson with an understanding of 'going at a medium pace', compared to hard or easy. The Perceived Intensity Assessment Task seemed to help children understand what it felt like for a person to keep going a long time.

It appeared from the other assessments--Homer Booklet and Video Tape Advertisement--that youngsters in the "inexperienced" group picked up on the words moderate or medium level from the Perceived Intensity Level Assessment Task and included that information in their product assessments. For this reason also, it is believed that this activity might better serve as an instructional unit rather than an actual performance assessment.

In the attempt to determine if, or how, scores can be used to make judgments about student learning the same three Content Experts evaluated the Perceived Intensity Level Station Assessment Task. All student results were available for perusal. Furthermore, all teacher and student task directions from beginning to end and teacher and student checklists were also available for content experts to examine. Again, the three content experts were asked to rate the assessment based on the thirteen content validity questions using the same scoring rubric as the Homer Booklet Assessment Task. See chapter 4, Figure 5 for Content Validity Rubric (Homer Task),

Content Expert Team

The average overall rubric score for all the questions by the three outside content experts was 1.86 out of a possible score of 2. Therefore, the Content Experts agreed that the final Perceived Intensity Level Assessment Task design met nearly all the criteria for high content validity.

One important note, Content Experts had some difficulty understanding that this assessment had a different purpose. The purpose was solely related to children's understanding of perceived intensity level and not the other three guidelines. The assessment was meant to measure children's ability to apply their understanding by coding themselves, or coding themselves and a partner. In this way, it was different from the Homer Booklet Assessment and Video Tape Advertisement Assessment. Once this was clear, the Content Experts were able to rate the task using the scoring rubric. Additionally, Content Experts noted that the validity scoring rubric may need to be re-written for this task, as some of the questions pertaining to rubrics and scoring criteria do not necessarily apply to this particular assessment task.

Calvin's insights indicate that "this would be a good lesson and self-assessment and use the video or booklet assessment after this lesson. This is a

better 'teaching' assessment. Videos and booklets are better 'measuring' assessments." As previously mentioned, the teacher and researcher agreed this might be the case for this assessment. The other Content Experts nodded in agreement, but also thought the Perceived Intensity Level Assessment could be refined and become a very effective performance assessment prior to fitness testing to help youngsters understand pacing and moderate intensity levels. All agreed pacing is a very difficult concept for youngsters at this age to understand and especially apply.

Moderate intensity was a new concept for these youngsters. It makes sense that children would think hard is better. They are taught to always try hard, do your best, work hard. So, being encouraged to go at a medium or moderate pace or intensity level is indeed a new insight.

Meaningfulness Question

When students were interviewed it became clear that the word meaningfulness did not mean much to them. Enjoyment, however, was a word children understood and could relate to therefore enjoyment was used as an indicator of meaningfulness to the youngsters. Furthermore, the last student reflection question asked students if they thought the information they were learning was important to know. The fact that students perceived the information as important to know was also interpreted by the researcher as one indicator of meaningfulness to the students.

Results and Discussion

Meaningfulness (Student Self-Reflection Question # 1 & #2)

Overall, most students (N=51) enjoyed the Perceived Intensity Level Assessment Task. Three found it okay eight did not enjoy it. One primary reason for enjoyment appeared to be task specific. In other words, students wrote that a

specific activity within the station was the reason for overall enjoyment. For example some wrote: "skating", "throwing the ball" and "running". The fact that the PIL Assessment allowed students to be active during physical education class time, or that it was "fun" or "cool" was also a popular response. Some students enjoyed the "workout" or "exercise" and "found it challenging for their bodies". Some indicated social reasons, such as, they enjoyed working with others, or enjoyed the music during the task. For example, Nile wrote: "yes I like the music".

Meaningfulness (Student Self-Reflection Question # 3)

Reasons written by the eight students who did not enjoy the assessment were that it made them sweaty or "burn up"; it was too tiring; or they did not enjoy a specific activity within the station task. A student in the "inexperienced" group wrote: "No! It is tyering." Specifically, what did you not like? "It is hard and tyering." This student's words seem to capture the insights of those few students who did not enjoy the activity. These comments seem to confirm what other researcher's and teacher's have found. That is, children do not necessarily like vigorous activity. Young children are more active for more minutes each day than adults, but it appears children may prefer lower intensity levels or brief spurts of intense activity shorter than one minute in length.

When students responded to what they did not like about the task, most responded "nothing", even if they indicated overall enjoyment. Some students, however, gave specific reasons for what they did not like. For example, a student in the "experienced" group, Kris (SC4), said: "I enjoyed it because I got to do fun things, I didn't like that I had to burn up." A student in the "inexperienced group, Ashley (SS4) wrote: "I enjoyed working with my friends. I did not enjoy sweating." When asked to respond to what changes could be made to the task, two children wrote to "make it more fun", but did not offer suggestions for how we might better

accomplish that. Two other students indicated that the twenty-second interval for coding and rotating was not long enough.

Interviews confirmed children's insights into this task and also revealed that some students learned from this task. One question to some children was: "Which assessment, of the three, helped you learn the most?" Some responded "the station task." For example, Ayn (SC4) said: "I thought the, um, stations, or whatever they were. The stations because you had to know if you were going at a medium, hard, or low level because you had a partner and stuff and you had to watch them as well as what you were doing at the same time and so you learned what moderate was for yourself, and hard, and low, and what it was for other people." Gen (SS4) said: "The stations, because when I was doing the stations I, um, kind of got tired afterwards and I was kind of feeling sick and I wasn't doing good. I figured that I should be more healthier."

Meaningfulness (Student Self-Reflection Question # 7)

Another question in the student reflection for all the assessments asked students if they thought knowing this information was important, why or why not. Again, all students responded that it is important to know about physical activity, and how to be healthy for a lifetime. In fact, this response was true of all three assessment tasks. Unlike the research that shows knowledge of health benefits is not important to youngsters, Gottlieb and Chen (1985) found that knowledge about "how to" was significant for youngsters in a weight loss program. This study seems to confirm those findings for this particular group of fourth and fifth grade youngsters not in a weight loss program.

Feasibility Question

Results and Discussion

The teacher found the Perceived Intensity Level Assessment Task to be a quick and easy visual check. In fact, in her interview, Patty stated that she liked this assessment the best of the three. She also commented that the format is more like what she is used to doing with the children, as she uses skill checklists often when assessing her students. Students are also very familiar with station activities. She liked this format. In her own words:

Patty: The station one I liked. That is more in the format that I'm used to using anyway. Um, I use stations a lot and it allows the kids to practice several different things during the class period and it allows the kids who are more 'ept' at physical ability to go ahead and experiment what they know. And it allows the kids to work that are at low level or still beginning level with a skill to have practice to improve. So, it gives them a little bit more freedom instead of being channeled into please do it this way--the whole class together...

Researcher: Uh hum

Patty: And then stations you know allows me, okay I'm going to be here watching, and I tell them the cues that I'm looking for...so evaluating the student's hard level, moderate level, or easy level it was really easy. I thought that was the easiest for me to do, and then with the additional input the kids were doing where they were rating themselves, you know, that gave me an idea of how much of an understanding where do we need to talk about...so, you know, the station one I liked and, um, rating a partner I like that one. I thought that was pretty neat where they were watching somebody else...And like I said, they were used to the station aspect...giving them twenty seconds to do their paper work and move is plenty of time. Um, the ones that were having a little bit of difficulty getting over there (changing stations on time) were the ones that were really trying to think about it and still were trying to grasp the concept how hard did I work?

Regarding the task coding sheet for the students Patty thought the last circle task score sheet was the easiest and best for the students. (See Figure 14 for details).

Patty: I liked the one that had the part with the circles. I think the circles for them was real, real clear...

Researcher: Um, do you think hard, medium, easy is a better way thing for them to write or the most of the time, some of the time, none of the time?

Patty: think most of the time, some of the time is kind of confusing. I think that the hard, moderate, and easy is a much easier comment.

Students also found the circle scoring sheet to be the easiest coding method as was previously mentioned. Furthermore, content experts agreed this was a very feasible assessment. Lee said: "This is very feasible to do in regular physical education class, and the activity is enjoyable for the kids." Finally, in a written evaluation of the assessment Patty wrote:

The station assessment is a quick and easy visual check. This type of assessment can be used for any skill check and can also be used in conjunction with written assessments. This would determine if a student could show correct form in a skill or whatever the area of content that is being assessed.

Conclusion & Interpretation of Results

The Perceived Intensity Level Assessment Task did not discriminate between youngsters that were taught and youngsters that were not taught the content of moderate intensity level during physical activity. It appears that a very high percentage of the students were able to match their coded intensity level with one or both raters. This might mean, that regardless of instruction, children at this age tend to understand how hard or easy their body is working during a short time interval.

If the purpose of the performance assessment had been to see how many children could continuously work at a moderate intensity level, results would probably indicate that many would not be able to. However, the purpose was to measure their understanding of intensity level. It was believed that understanding their own intensity level is important for youngsters. Identifying how hard or easy

their body is working might be helpful for them in knowing if they are meeting the general guidelines for health purposes. Ultimately, these children might find and participate in activities that help promote appropriate intensity levels for health, which is a more relaxed guideline than for physical fitness. Determining any carryover of understanding would involve another study. Following fourth and fifth grade students, who have been taught and not taught these concepts, into their adolescent years would be an interesting and recommended study for the future.

The teacher and Content Experts found the Perceived Intensity Level Assessment Task to have high content validity. This finding, however, may be irrelevant as the assessment did not discriminate between youngsters who were taught and those not taught. Most students were able to match their coding with adult raters. Students did find the assessment task meaningful and enjoyable, and the teacher and Content Experts found it to be very feasible in a regular physical education setting.

Again, this task might better serve as an instructional unit and self-check prior to fitness testing, versus the purpose for which it was developed. Nevertheless, this researcher believes that with further refinement and modifications, a task similar to this might become an alternative assessment to the now "inappropriate" physical fitness tests currently in use nationwide. While many students do not enjoy the mile run, sit-ups, pull-ups, and sit and reach, most students enjoyed this very active assessment. This PIL assessment appears to be more "authentic" and "play like" for youngsters at this age than the traditional fitness tests.

Developing alternative assessments that measure physical activity or understanding of physical activity is important. Researchers and teachers should work together with children to ensure we, as physical educators, are measuring

appropriate content and performances that might contribute to healthy lifestyles for all people in years to come.

CHAPTER 7

Summary and Conclusion

This chapter includes the final survey, and conclusions about the use of the three alternative assessments which were designed, piloted, and evaluated for this study. The final survey was given post facto to all students to gain their insights into which assessment task of the three students preferred overall. Results of the final survey are included in this chapter because they do not fit into any of the previous results chapters for the three individual assessment tasks.

Final Survey

On a final questionnaire called "Your Favorite Activity" students' were asked which of the three assessments they liked the best. Two of the alternative assessments measured the same content, but in different mediums. One (Homer Booklet Assessment Task) was in a written medium, the other (Video Tape Advertisement Assessment) was a verbal and "acting" medium. Interestingly, regardless of instruction--"experienced" or "inexperienced" with the content--the majority of all students from all four classes in this study liked the Video Tape Advertisement Assessment the best. The teacher, however, found the video to be the least feasible to administer and the most time consuming.

Students were asked: "If a friend at another school could only do one of the tasks we did (Stations, Homer Booklet, or Video Tape Advertisement) which one would you pick for your friend to do?" The question was asked this way because it was assumed by the researcher, that students would pick for a friend the task they liked the best.

In fact forty students preferred the Video Tape Advertisement, sixteen preferred the Perceived Intensity Level Station Assessment, and only three choose

the Homer Booklet Assessment Task. In addition, two students responded that it would depend on the friend and what that friend liked. Ultimately, the final survey revealed that most students, although liking all the tasks, particularly enjoyed the Video Tape Advertisement Assessment Task.

Summary

The purpose of this study was to design, pilot, and evaluate three alternative assessments that measured fourth and fifth grade students understanding of revised guidelines for physical activity as related to health. These guidelines were drawn from the Surgeon General's Report (USDHHS, 1996). Understanding was defined as the ability to apply facts, concepts, and skills in new situations. Those new situations included: 1) designing a booklet for a peer named Homer; 2) creating a Video Tape Advertisement for students at another school; and 3) coding the perceived intensity level for one minute intervals at an active aerobic type station (Perceived Intensity Level Assessment Task).

Four research questions guided the evaluation of the assessment tasks. They were: 1. Do students who have been taught the physical activity guidelines score differently on the alternative assessment than those who have not been taught the guidelines? 2. Do content experts agree that scores can be used to describe what students have learned (content validity)? 3. Do students find the assessment task worthwhile, enjoyable, and meaningful? 4. Is the alternative assessment prototype "feasible" for a teacher to administer in a regular physical education setting?

Results from this study indicate that the Homer Booklet Assessment Task discriminated between those who were taught and those who were not taught. Furthermore, both the Content Experts and the teacher found this assessment to have high content validity and found it to be feasible to use in a regular physical education setting. The students wrote that they enjoyed the Homer Booklet

Assessment Task on their self-reflections. However, students found the Homer Booklet the least enjoyable of the three assessment tasks.

In contrast to the Homer Booklet Assessment Task, students enjoyed the Video Tape Advertisement Task best. Although this assessment task did discriminate in content between those who were taught and not taught there were a couple of problems with this assessment. There were specific problems with style scores in that scores only discriminated with fifth grade students--but not fourth. Furthermore, feasibility--specifically the length of time needed to complete this assessment--seems to be problematic with this assessment. Despite the problems with the Video Tape Advertisement Assessment Task, content experts found it to have high content validity.

Finally, the Perceived Intensity Level Assessment Task did not discriminate between students who were taught and not taught. Content Experts found it to have high content validity, however, they thought it would be a better instructional task than an assessment task. Most students found the Perceived Intensity Level Assessment task as enjoyable and content Experts and the teacher found it to be very feasible to administer in a regular physical education setting.

Conclusions

Based on the results of this study, the Homer Booklet Task is ready to use in other settings. The Video Tape Advertisement Assessment, however, needs further revisions before it can be used, and the Perceived Intensity Level Assessment Task is not ready for use as an assessment in other settings since it did not discriminate between students who were taught and not taught the content.

If a teacher wants to quickly measure what students have learned about specific content, guidelines, or cues, designing and implementing a Homer-type assessment task might be the best choice. If a teacher wants to assess performance,

video tape assessments might be a good medium to heighten the interest, enjoyment of the process, and, ultimately, the meaningfulness for students. Teachers have to make tough decisions and tough choices to best meet the needs of their students in their programs. Alternative assessments such as those designed for this study are time-consuming. Therefore, care should be taken to make sure the assessment is measuring worthy and important subject matter.

Suggestions for further Research

The three assessments designed for this study should be "played with" and evaluated in numerous contexts, in numerous settings, with numerous youngsters, before final judgments can be made. The Perceived Intensity Level Assessment Task might be tried against objective measures to determine reliability. Furthermore, other performance assessments measuring physical activity versus physical fitness should be developed.

In addition, as physical activity guidelines and cognitive understanding are not the only important content in physical education, other product and performance assessments should be developed, piloted, and evaluated. Teachers do not always have the time to develop labor-intensive alternative assessments and researchers need student examples to develop appropriate rubrics and tasks. Therefore, researchers and teachers should work together in the development of meaningful performance assessments.

Major Contributions of this Study

Outcomes related to health, quality of life, and longevity of life were viewed as important to the students, Review Team, and all Content Experts involved in this study. Assessments measuring students' understanding of these concepts are important to the field of physical education. Furthermore, the nation is certainly concerned with health, as evidenced by the goals 2000 project (USDHHS, 1991), the Surgeon General's Report (USDHHS, 1996), and concern with

national health care costs in general. Developing three alternative assessment tasks that measured understanding of this meaningful content was needed.

Assessment tasks, such as the Homer Booklet and Video Tape Advertisement can identify learning in physical education that might further the promotion of physical education programs across the nation. Physical education is often viewed as simply playing games. Unfortunately, some programs are just that. Chapter 2 began with a statement that physical activity is the essence, the foundational basic nature, the root of the tree. This study was one example of important content (healthy roots) valued, taught and assessed in physical education.

Implications of Results

As accountability for programs increases, need for evidence of learning is more and more necessary. Even within the limitations of this study (one setting, few students, one teacher), clear contributions to the field of physical education and specifically, to alternative assessment, are evident. If teachers can use motivating and meaningful alternative assessments to show evidence of what their students know, or can do, as a result of instruction, physical education programs would have powerful evidence of student learning. With this type of evidence of learning, physical education programs might move away from the marginalized position they occupy towards a more vital role in school settings.

Two of the assessments in this study measured students' cognitive understanding of critical content in physical education. Since it can be argued that knowledge comes before behavior, measuring student understanding becomes important. Using alternative, and even more "authentic" type assessments to measure student understanding is meaningful to youngsters. Furthermore, using alternative-like assessments is the direction other subject areas (History, writing, whole language, science, mathematics) are moving toward nationwide.

Physical fitness tests typically do not "show off" what students know and can do as a result of physical education programs, especially when youngsters only participate in two or three day a week programs. Furthermore, there is some evidence that children do not enjoy and try their best on tasks like the mile-run and pull-ups, for example. Fitness tests are very task specific and if children do not train for them they will not score well. Being physically active on a daily basis is healthy, however, and much different from fitness training.

Physical activity has been shown to be important to a healthy lifestyle. This study has shown that most instructed youngsters were able to write or tell someone else how to be physically active for a lifetime. They were also able to write or tell someone to select physical activities that are "fun for you." Understanding how much time, how many days, and what intensity level to be active daily according to the revised guidelines for health, is important content for physical education. Understanding the importance of selecting activities that are "fun for you" is also important content for physical education.

In addition, closely matching assessment and instruction are important to document student learning. For this study, the Homer Booklet Assessment Task and Video Tape Assessment Task were closely related to the instruction. Physical fitness tests do not measure cognitive understanding and would not reveal this important student understanding.

The three alternative assessments designed for this study are simply a beginning into what future assessment and accountability might look like in physical education. This study was an initial look at the kaleidoscope of assessment possibilities.

REFERENCES

- American College of Sports Medicine [ACSM]. (1978). The recommended quantity and quality of exercise for developing and maintaining fitness in healthy adults. Medicine and Science in Sports and Exercise, 22, 265-274.
- Ainsworth, B., Haskell, W., Leon, A., Jacobs, D., Montoye, H., Sallis, J., & Paffenbarger, R. (1993). Compendium of physical activities: classification of energy costs of human physical activities. Medicine and Science in Sports and Exercise, 25, 71-80.
- Allensworth, D. (1996). Cardiovascular objectives for youth in Healthy People 2000: Update on the status of risk factors. Journal of Health Education, 27 (Suppl.), 17-21.
- Baranowski, T., Bouchard, C., Bar-Or, O., Bricker, T., Health, G., Kim, S.Y.S., Malina, R., Obarzanek, E., Pate, R., Strong, W., Truman, B., & Washington, R. (1992). Assessment, prevalence, and cardiovascular benefits of physical activity and fitness in youth. Medicine and Science in Sports and Exercise, 24, (suppl), S237-247.
- Baranowski, T., Dworkin, R. J., Cieslik, C. J., Hooks, P., Clearman, D. R., Ray, L., Dunn, J. K., & Nader, P. R. (1984). Reliability and validity of self report of aerobic activity: Family Health Project. Research Quarterly for Exercise and Sport, 55, 309-317.
- Blair, S. N. (1992). Are American youth fit? The need for better data. Research Quarterly for Exercise and Sport, 63, 120-123.
- Blair, S. N., Clark, D. G., Cureton, K. J., & Powell, K. E. (1989). Exercise and fitness in childhood: Implications for a lifetime of health. In C. V. Gisolfi & D. R. Lamb (Eds.), Perspectives in exercise science and sports medicine, Vol. 2: Youth, exercise, and sport (pp. 401-430). Indianapolis, IN: Benchmark Press.
- Bloom, B. S., (1956). Taxonomy of educational objectives, Vol. 1: Cognitive domain. White Plains, NY: Longman.
- Blumberg, F., Epstein, M, MacDonald, W., & Mullis, I. (1986). A pilot study of higher-order thinking skills assessment techniques in science and mathematics: Final Report-Part I. Princeton, NJ: National Assessment of Educational Progress, National Science Foundation, Office of Educational Research and Improvement - Grant no. NIE-G-84-2006-P4.
- Bogdan, R. C., & Biklen, S. K. (1982). Qualitative research for education: An introduction to theory and methods, Boston: Allyn and Bacon, Inc.

Borra, S. T., Schwartz, N. E., Spain, C. G., & Natchipolsky, M. M., (1995). Food, physical activity, and fun: Inspiring America's kids to more healthful lifestyles. Journal of the American Dietetic Association, 7, 816-818.

Brill, P. A., Burkhalter, H. E., Kohl, H. W., Blair, S. N., & Goodyear, N. N. (1989). The impact of previous athleticism on exercise habits, physical fitness, and coronary heart disease risk factors in middle-aged men. Research Quarterly for Exercise and Sport, 60, 209-215.

Brown, A. L. (1994). The advancement of learning. Educational Researcher, 23, 4-12.

Bruner, J. (1979). On knowing: Essays for the left hand (Expanded Edition). London, England: Belknap Press of Harvard University Press.

Brunner, B. C. (1969). Personality and motivation factors influencing adult participation in vigorous physical activity. Research Quarterly, 40, 464-469.

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), 126-130.

Corbin, C. B., & Pangrazi, R. P. (1992). Are American children and youth fit? Research Quarterly for Exercise and Sport, 63, 96-106.

Cronbach, L. J. (1983). Designing evaluations of educational and social programs. In G. F. Madaus, M. S. Scriven & D. L. Stufflebeam (Eds.), Evaluation models. (pp. 357-379). Boston, MA: Kluwer-Nijhoff Publishing.

Dewey, J., & Dewey, E. (1915/1962). Schools of tomorrow. New York: E. P. Dutton & Co., Inc.

Driver, R., Asoko, H., Leach, J., Mortimer, E., & Scott, P. (1994). Constructing scientific knowledge in the classroom. Educational Researcher, 23, 5-12.

Educators in Connecticut's Pomperaug Regional School District 15. (1996). A teacher's guide to performance-based learning and assessment. Alexandria, VA: Association for Supervision and Curriculum Development.

Faucette, N., Sallis, J. F., McKenzie, T. L., Alcaraz, J., Kolody, B., & Nugent, P. (1995). Comparison of fourth grade students' out-of-school physical activity levels and choices by gender: Project SPARK. Journal of Health Education, 26 (Suppl.), S82-90.

Fox, K., & Biddle, S. (1992). The use of fitness tests: Educational and psychological considerations. Journal of Physical Education, Recreation and Dance, 59 (2), 47-53.

Fox K., & Biddle, S. (1988). This child's perspective in physical education: Pt. 2. Children's participation motives. British Journal of Physical Education, 19 (2), 79-82.

Frederiksen, N. (1984). The real test bias. American Psychologist, 39, 193-202 (p.199).

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

Freedson, P. S. (1991). Electronic motion sensors and heart rate as measures of physical activity in children. Journal of School Health, 61, 220-223.

Freedson, P. S. (1989). Field monitoring of physical activity in children. Pediatric Exercise Science, 1, 8-18.

Friedman, W. J., (1982). (Ed.) The developmental psychology of time. New York: Academic Press.

Gardner, H. (1991). The unschooled mind. New York: Basic Books.

Gardner, H. (1983). Frames of mind. New York: Basic Books.

Getty, V. M. (1995). Food, physical activity, & fun, what kids think: Key findings of gallop survey. University Park, PA: Penn State Nutrition Center. Document Number 121011425.

Gottlieb, N. H., & Chen, M. S. (1985). Sociocultural correlates of childhood sporting activities: Their implications for heart health. Social Science in Medicine, 21 (5), 533-539.

Graham, G. (September 15, 1996). Personal communication.

Graham, G., Hopple, C., Manross, M., & Sitzman, T. (1993). Novice and expert children's physical education teachers: Insights into their situational decision-making. Journal of Teaching Physical Education, 12, 197-214.

Gutin, B., Manos, T., & Strong, W. (1992). Defining health and fitness: First step toward establishing children's fitness standards. Research Quarterly for Exercise and Sport, 63, 128-132.

Hales, D. (1991). Your health. Redwood City, CA: Benjamin Cummings Publishing Co.

Hansen, N., Landsmann, L., & Monismith, S. (1996). Schools play a role in fighting heart disease. Journal of Health Education, 27 (Suppl.), 8-11.

Herman, J. L., Aschbacher, P. R., & Winters, L. (1992). A practical guide to alternative assessment. Alexandria, VA: Association for Supervision and Curriculum Development.

Hopper, C., Gruber, M., Munoz, K., & MacConnie, S. (1996). School-based cardiovascular exercise and nutrition programs with parent participation. Journal of Health Education, 27 (Suppl.), 32-39.

Hopple, C. (1995). What children think, feel, and know about physical fitness testing. Journal of Teaching in Physical Education, 14, 408-417.

Hopple, C. (1994). What children think, feel, and know about physical fitness testing. Unpublished master's thesis, Virginia Polytechnic Institute and State University, Blacksburg.

Janz, K. F., Witt, J., & Mahoney, L. T. (1995). The stability of children's physical activity as measured by accelerometry and self-report. Medicine and Science in Sports and Exercise, 27, 1326-1332.

Kanfer, F. H., & Goldstein, A. P. (1980). Helping people change. New York: Pergamon.

Krahenbuhl, G. S., Skinner, J. S., & Kohrt, W. M. (1985). Developmental aspects of maximal aerobic power in children. In R. L. Terjung (Ed.), Exercise science and sports research (pp. 503-538). New York: Macmillan.

Lambert, L. (1996). Goals and outcomes. In S. J. Silverman & C. D. Ennis (Eds.), Student learning in physical education (pp. 150-169). Champaign, IL: Human Kinetics.

Lambert, L. (October 11, 1996). Personal communication.

LaPorte, R. E., Montoye, H. J., & Caspersen, C. J. (1985). Assessment of physical activity in epidemiologic research: Problems and prospects. Public Health Reports, 100, 131-146.

Locke, L. (1989). Qualitative research as a form of scientific inquiry in sport and physical education. Research Quarterly for Exercise and Sport, 60 (1), 1-20.

Malina, R. M. (1994). Physical activity: Relationship to growth, maturation, and physical fitness. In C. Bouchard, R. Shephard, & T. Stephens (Eds.), Physical activity, fitness, and health (pp. 918-930). Champaign, IL: Human Kinetics.

Manross, M. (1994). What children think, feel, and know about the overhand throw. Unpublished masters thesis, Virginia Polytechnic Institute and State University, Blacksburg.

Mathematics Sciences Education Board (1993). Measuring up: Prototypes for mathematics assessment. Washington DC: National Academy Press.

Marzano, R. J. (1992). A different kind of classroom: Teaching with dimensions of learning. Alexandria, VA: Association for Supervision and Curriculum Development Publications.

McKenzie, T. (September 3, 1996) Personal correspondence. Electronic mail.

McKenzie, T., Feldman, H., Woods, S., Romero, D., Dahlstrom, V., Stone, E., Strikmiller, P., Williston, J., & Harsha, D. (1995) Student activity levels and lesson context during third grade physical education. Research Quarterly for Exercise and Sport, 66, 184-193.

McKenzie, T., Nader, P. R., Strikmiller, P. K., Yang, M., Stone, E. J., Perry, C. L., Taylor, W. C., Epping, J. N., Feldman, H. A., Luepker, R. V., & Kelder, S. H. (1996). School physical education: Effect of the child and adolescent trial for cardiovascular health. Preventive Medicine, 25, 423-431.

McKenzie, T., & 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 (pp. 223-246). Champaign, IL: Human Kinetics.

McTighe, J., & Ferrara, S. (1994). Assessing Learning in the classroom: A report from professional standards and practice. Washington, DC: National Education Association.

Meyer, C. A. (1992). What's the difference between authentic and performance assessment? Educational Leadership, 49, 39-40.

Myers, L., Strikmiller, P. K., Webber, L. S., & Berenson, G. S. (1996). Physical and sedentary activity in school children grades 5 - 8: Bogalusa heart study. Medicine and Science in Sports and Exercise, 28, 852-859.

National Association for Sport and Physical Education. (1992). Outcomes of quality physical education. Reston, VA: American Alliance for Health, Physical Education, Recreation and Dance.

National Association for Sport and Physical Education (NASPE). (1995). Moving into the future: A guide to content and assessment. St. Louis, MO: Mosby

Neill, D. M., & Medina, N. J. (1989). Standardized testing: Harmful to educational health. Phi Delta Kappan, 70, 688-697.

New Standards Project (1992). Aquarium problem and teacher guidelines. Learning Research and Development Center University of Pittsburgh, PA: University of Pittsburgh.

O'Connell, J. K., Price, J. H., Roberts, S. M., Jurs, S. G., & McKinley, R. (1985). Utilizing the health belief model to predict dieting and exercising behavior of obese and nonobese adolescents. Health Education Quarterly, 12, 343-351.

O'Neil, J. (1992). Putting performance assessment to the test. Educational Leadership, 49, 14-19.

Pate, R. R., & Blair, S. N. (1978). Exercise and the prevention of atherosclerosis: Pediatric implications. In W. Strong (Ed.), Pediatric aspects of atherosclerosis (pp. 251-286). New York: Grune & Stratton.

Pate, R. R., Dowda, M., & Ross, J. G. (1990). Associations between physical activity and physical fitness in American children. American Journal of Diseases of Children, 144, 1123-1129.

Pate, R. R., Long, B. J., & Heath, G. (1994). Descriptive epidemiology of physical activity in adolescents. Pediatric Exercise Science, 6, 434-447.

Pate, R. R., Pratt, M., Blair, S. N., Haskell, W. L., Macera, C. A., Bouchard, C., Buchner, D., Ettinger, W., Heath, G. W., King, A. C., Kriska, A., Leon, A. S., Marcus, B. H., Morris, J. L., Paffenbarger, R. S., Jr., Patrick, K., Pollock, M. L., Rippe, J. M., Sallis, J., & Wilmore, J. H. (1995). Physical activity and public health: A recommendation from the Centers for Disease Control and Prevention and the American College of Sports Medicine. Journal of the American Medical Association, 273 (5), 402-407.

Patton, M. Q. (1990). Qualitative evaluation and research methods (2nd. ed.). Newbury Park, CA: Sage Publications.

Perrone, V. (1991). Expanding student assessment. Alexandria VA: Association for Supervision and Curriculum Development.

Powell, K. E., & Dysinger, W. (1987). Childhood participation in organized sports and physical education as precursors of adult physical activity. American Journal of Preventative Medicine, 3, 276-281.

Resnick, L. B. (1987). Learning in school and out: The 1987 Presidential Address. Educational Researcher, 16, 3-20.

Resnick, L. B., & Klopfer, L. E. (Eds.). (1989). Toward the thinking curriculum: Current cognitive research. 1989 Yearbook of the Association for Supervision and Curriculum Development. Alexandria VA: Association for Supervision and Curriculum Development.

Resnick, L. B., & Resnick, D. P. (1985). Standards, curriculum, and performance: A historical and comparative perspective. Educational Researcher, 14, 5-20.

Ross, J. G., & Gilbert, G. G. (1985). The national children youth fitness study: A summary of findings. Journal of Physical Education, Recreation, & 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, 56(1), 51-56.

Safrit, M. J., & Looney, M. A. (1992). Should the punishment fit the crime? A measurement dilemma. Research Quarterly for Exercise and Sport, 63, 124-127.

Sallis, J. (September 3, 1996). Personal correspondence. Electronic mail.

Sallis, J. F. (1993). Epidemiology of physical activity and fitness in children and adolescents. Critical Reviews in Food Science and Nutrition, 33 (4/5): 403-408.

Sallis, J. F., Berry, C. C., Broyles, S. L., McKenzie, T., & Nader, P. R. (1995). Variability and tracking of physical activity over 2 yr. in young children. Medicine and Science in Sports and Exercise, 27 (7), 1042-1049.

Sallis, J. F., Buono, M. J., Roby, J. J., Carlson, D., & Nelson, J. A. (1990). The caltrac accelerometer as a physical activity monitor for school-age children. Medicine and Science in Sports and Exercise, 22, 698-703.

Sallis, J. F., Buono, M. J., Roby, J. J., Micale, F. G., & Nelson, J. (1993). 7-Day recall and other physical activity self-reports in children and adolescents. Medical and Science in Sports and Exercise, 25, 99-107.

Sallis, J. F., Bruce, G., Stone, C. B., Simons-Morton, E. J., Corbin, C. G., Epstein, L. H., Faucette, N., Iannotti, R. J., Killen, J. D., Klesges, R. C., Petray, C. K., Rowland, T. W., & Taylor, W. C. (1992). Determinants of physical activity and interventions in youth. Medicine and Science in Sports and Exercise, 24, (6), S248-S257.

Sallis, J. F., Condon, S. A., Goggin, K. J., Roby, J. J., Kolody, B., & Alcaraz, J. E. (1993). The development of self-administered physical activity surveys for 4th grade students. Research Quarterly for Exercise and Sport, *64*, 25-31.

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

Sallis, J. F., McKenzie, T., & Alcaraz, J. E. (1993). Habitual physical activity and health related physical fitness in fourth-grade children. American Journal of Diseases of Children, *147*, 890-896.

Sallis, J. F., Strikmiller, P. K., Harsha, D. W., Feldman, H. A., Ehlinger, S., Stone, E. J., Williston, J. & Woods, S. (1996). Validation of interviewer-and-self administered physical activity checklists for fifth grade students. Medicine and Science in Sports and Exercise, *28*, 840-851.

Sanders, S. (1993). Kindergarten children's initial experiences in physical education. Unpublished doctoral dissertation, Virginia Polytechnic and State University, Blacksburg.

Saris, W. H. (1986). Habitual physical activity in children: Methodology and findings in health and disease. Medicine and Science in Sports and Exercise, *18*, 253-263.

Shavelson, R. J., & Baxter, G. P. (1992). What we've learned about assessing hands-on science. Educational Leadership, *49*, 20-25.

Shepard, L. (1995). Using assessment to improve learning. Educational Researcher, *52*, 38-43.

Simons-Morton, B. G., Taylor, W. C., & Huang, I. W. (1994). Validity of the physical activity interview and caltrac with preadolescent children. Research Quarterly for Exercise and Sport, *65*, 84-88.

Simons-Morton, B. G., Parcel, G. S., O'Hara, N. M., Blair, S. N., & Pate, R. R. (1988). Health-related physical fitness in childhood: Status and recommendations. Annual Review of Public Health, *9*, 403-425.

Simons-Morton, B. G., O'Hara, N. M., Simons-Morton, D. G., & Parcel, G. S. (1987). Children and fitness: A public health perspective. Research Quarterly for Exercise and Sports, *58*, 295-302.

Sizer, T. (1991). No pain, no gain. Educational Leadership, *48*, 32-34.

Sleap, M., & Warburton, P. (1992). Physical activity levels of 5-11-year-old children in England as determined by continuous observation. Research Quarterly for Exercise and Sport, 63, 238-245.

Stecher, B. M., & Klein, S. P. (1996). Performance assessments in science: Hands-on tasks and scoring guides. RAND Institute on Education and Testing, National Science Foundation.

Stiggins, R. J. (1994). Student centered classroom assessment. New York, NY: Macmillan College Publishing Company.

Strong, W. B. (1990). Physical activity and children. Circulation, 81, 1697-1701.

Strong, W. B., & Wilmore, J. H. (1988). Unfit kids: An office-based approach to physical fitness. Contemporary Pediatrics, 5 (4), 33-48.

Strouse, J., & Fabes, R. (1985). Formal versus informal sources of sex education: Competing forces in the sexual socialization of adolescents. Adolescence, 20, (78), 251-263.

Stucky-Ropp, R. C., & DiLorenzo, T. M. (1993). Determinants of exercise in children. Preventive Medicine, 22, 880-889.

Tinsley, B. J., Holtgrave, D. R., Reise, S. P., Erdley, C. & Cupp, R. G. (1995). Developmental status, gender, age, and self-reported decision-making influences on students' risky and preventive health behaviors. Health Education Quarterly, 22, 3244-3259.

Updyke, W. F. (1992). In search of relevant and credible physical fitness standards for children. Research Quarterly for Exercise and Sport, 63, 112-119.

U.S. Department of Health and Human Services (1996). Physical activity and health: A report of the Surgeon General. Atlanta, GA: Author.

U.S. Department of Health and Human Services (1991). Healthy people 2000: National health promotion and disease prevention objectives (DHHS Publication No 91-50212). Washington, DC: Author.

Wankel, L. M. (1988). Exercise adherence and leisure activity: Patterns of involvement and interventions to facilitate regular activity. In R. K. Dishman (Ed.), Exercise adherence: Its impact on public health (pp. 369-396). Champaign, IL: Human Kinetics.

Wiggins, G. (1993). Assessing student performance: Exploring the purpose and limits of testing. San Francisco, CA: Jossey-Bass Publishers.

Wiggins, G. (1995). The CLASS Conference on assessment reform. Raleigh, NC (September 14-16, 1995). Princeton, NJ: The Center on Learning Assessment, and School Structure (CLASS).

Wolf, D., Bixby, J., Glen, III, & Gardner, H. (1991). To use their minds well: Investigating new forms of student assessment. In G. Grant (Ed.), Review of Research in Education (pp. 31-44). Washington, DC: American Educational Research Association.

Wood, T. M. (1996). Evaluation and testing: The road less traveled. In S. J. Silverman & C. D. Ennis (Eds.), Student learning in physical education: (pp. 150-169). Champaign, IL: Human Kinetics.

Woods, J., Pate, R., & Burgess, M. (1992). Correlates to performance on field tests of muscular strength. Pediatric Exercise Science, 4, 302-311.

Appendix A
Informed Consent

January 21, 1997

Dear Student:

My name is Natalie Doering and I am a student and physical education teacher at Virginia Tech studying children's physical education. Before I came to Virginia Tech to study and teach, I was an elementary physical education teacher for six years in Orlando, Florida.

I am asking for your help with the study I am doing. The study will involve helping me design an alternative to the current fitness test. You might be asked to respond in writing to what you do during the regular physical education class, do booklets and other tasks related to physical activity. You may also be asked to be interviewed by me. The regular p.e. teacher will always be present. I will not ask you to do anything you would not ordinarily do in the classroom or in the p.e. class. At any time you can decide not to participate in the study. If I share the results with others I will use pseudonyms (pretend names) for the school, the teacher, and all of the children in the study.

I hope you enjoy the lessons and assessments. I am looking forward working with you.

Sincerely,

Natalie Doering

Please retain this part of the paper for any future needs.

Should I have any questions about this research or its conduct, I may contact:

Natalie Doering (Investigator)

Phone: 961-5645

George Graham (Faculty Advisor)

Phone: 231-7545

Tom Hurd (Chair of I. R. B.)

Phone: 231-5281

Student Informed Consent of Investigative Project

If you agree to participate in this project please sign your name below and return this part to your teacher by **Wednesday, January 22, 1997**. Thank you.

(Your signature/student)

(Your Classroom Teacher's Name)

Parent Informed Consent

January 21, 1997

Dear Parent or Guardian:

My name is Natalie Doering and I am a student and physical education teacher at Virginia Tech studying children's physical education. Before I came to Virginia Tech to study and teach, I was an elementary physical education teacher for six years in Orlando, Florida. I am interested in designing an alternative assessment so that one day, fourth and fifth grade students might be able to do this assessment instead of the physical fitness tests.

I am asking for your child's help with this study. Your child might be asked to respond in writing to what he/she does during the regular physical education class, do booklets and other tasks related to physical activity. She/he may also be asked to be interviewed by me. The regular p.e. teacher will always be present. I will not ask your child to do anything he/she would not ordinarily do in the classroom or in the p.e. class. At any time you can decide for your child not to participate in the study and her/his work will not be included in the data. If I share the results with others I will use pseudonyms (pretend names) for the school, the teacher, and all of the children in the study.

I hope your child enjoys the lessons and assessments. If you have questions now, or at any time, please do not hesitate to call me. I can be reached at the university (xxx-xxxx) or at home (xxx-xxxx). I am looking forward to this project and working with your child.

Sincerely,

Natalie Doering

Parent or Guardian Informed Consent of Investigative Project

I hereby grant permission for my child to be a participant in the physical education project described above. Please return this part by **Friday, January 22, 1997**.

(Child's name)

(Classroom Teacher's name)

(Parent/Guardian Signature)

(Date)

APPENDIX B

Example of Alternative Assessment Tasks

Table B 1

Assessment Example 1

Many students your age do not know how to be physically active according to the new way you have learned. They do not know this information because it is so new. They also do not know why it is important to be physically active. You can help a friend learn this information by designing a booklet that they can read. The booklet should be interesting and fun for them to read. Include the information your friend needs to know about how to be physically active. Explain why it is important for them to be physically active for a lifetime. You may include drawings and art work to help them understand.

RUBRIC

4 The student clearly explains the time guidelines. The student clearly explains the number of days needed to be active for health purposes. The student gives examples of how to break up the day to build activity into their daily lives. The student may break up the day in different ways that add to the right number of minutes. The student explains the various levels of intensity and gives the appropriate target level. They should include how the friend will know they are active at that level. The student clearly explains to the friend how to pick the activities and why they should base decisions on those reasons.

3 The student explains the time, day, and intensity guidelines. The student includes how to select activities. No reasons, or poor examples or explanations are used with the information. The booklet is interesting and fun for a friend to read.

2 The student does not include two or less of the major components for how to be physically active. Reasons may be given, but the information is lacking and therefore the friend may not be able to learn about the new ways to be healthy and active for a lifetime. The booklet may, or may not be interesting and fun to read.

1 The student includes less than three of the major components needed for how to be physically active to health reasons. The student does not explain why or give examples for the information they include. The booklet is not interesting and fun for a friend to read. Drawings and art work are not used to help the student want to read the booklet.

Appendix C
Homer Booklet Task
Phase 4

Homer does not know about the new Surgeon General's Report. He does not know this information because it is so new. This report explains how to gain health benefits through a lifetime of physical activity. In this class, you have learned how to be physically active to meet the standards for health benefits. You can help Homer learn this information by designing a booklet that he can read. The booklet should include all of the information Homer needs to be physically active for a lifetime. The booklet should also be interesting and enjoyable for him to read. Include drawings and art work to help Homer understand.

RUBRIC

Level 4: Surgeon General

- The student clearly and accurately explains the critical components (number of days, time, and intensity level) needed to obtain health benefits.
- The student clearly explains to the reader how to choose activities that are appropriate for them.
- The booklet should be interesting and enjoyable to read. The booklet should be publishable.

Level 3: Surgeon General's Aide

- The student includes all the necessary information--the same as above.
- The booklet may or may not be interesting and enjoyable for someone else to read.

Level 2: Apprentice Aide

- The student clearly and completely explains most of the important points, but, one or more important parts are missing. Therefore, the reader will not quite know how to be physically active for a lifetime to obtain health benefits.
- The booklet may, or may not be interesting and enjoyable to read.

Level 1: Novice Aide

- The critical components are incomplete, incorrect, or not the critical components learned during physical education class.

Level 0:

- The student chooses not to do the booklet.

APPENDIX D

Test Construction

This section elaborates on some specifics about how to construct and develop assessment tasks. Herman, Aschbacher, and Winters (1992) provide practical guides to the development of these types of assessment. The following table is taken directly from their book "A Practical Guide to Alternative Assessment" (p. 42).

Table D 1

Checklist for Task Description

Outcomes to be Measured	<ul style="list-style-type: none"> • Description of instructional goals • Eligible content/Topics • Rules/Process for selection
Assessment Administration Process	<ul style="list-style-type: none"> • Group/Individual roles • Materials/Equipment • Administration instructions • Help allowed • Time allowed
Actual Question/Problem/Prompt	<ul style="list-style-type: none"> • Format • Audience • Options available • Student directions
Scoring	<ul style="list-style-type: none"> • Rubric/Criteria • Scoring procedures • Use of scores

The following section, in outline format, comes from chapter three through five of the same book (Herman, Aschbacher, and Winters (1992, pp. 23 - 79). Many of these recommendations come from by Linn, Baker, & Dunbar, 1991 as cited in Herman, Aschbacher, & Winters, 1992).

1. DETERMINING PURPOSE

A. Setting primary instructional goals

B. Determining priority outcomes

cognitive, social, metacognitive, types of problems, what concepts and principles should students be able to apply?

C. Using available resources

- 1) NASPE Standards
- 2) Surgeon General Guidelines for physical activity
- 3) State and curriculum frameworks
- 4) Consulting with colleagues

D. Setting meaningful priorities:

- 1) How much time will it take?
- 2) How does skill relate to other complex cognitive, social, and affective skills?
- 3) How does skill relate to long-term school and curricular goals?
- 4) What is the intrinsic importance of skills
- 5) Are the desired skills teachable and attainable for students?

2. CHOOSING GOOD TASKS

- A. Does the task match specific instructional intentions?
- B. Does the task adequately represent the content and skills you expect students to attain?
- C. Does the task enable students to demonstrate their progress and capabilities?
- D. Does the assessment use authentic, real-world tasks?
- E. Does the task lend itself to an interdisciplinary approach?
- F. Can the task be structured to provide measures of several goals?

3. DESCRIBING THE ASSESSMENT TASK

- A. what outcomes are intended for the assessment?

- B. What are the eligible content/topics?
- C. What is the nature and format of questions to be posed to students?
- D. Is it group or individual work? Roles?
- E. What materials/equipment/resources will be available to students? Are there any specifications?
- F. What directions will be given to students?
- G. What administrative constraints are there? How much time is allowed? What is the order of tasks? How will student questions be answered? What help will be allowed?
- H. What scoring scheme and procedures will be used?

4. ENSURING THAT YOUR TASKS LEAD TO SOUND ASSESSMENTS

- A. Do the tasks match the important outcome goals?
- B. Do they pose enduring type problems students will continue to face in school and future lives?
- C. Are the tasks fair and free of bias?
- D. Will the tasks be credible to important constituencies?
- E. Will the tasks be meaningful and engaging to students so that they will be motivated to show their capabilities?
- F. Are the tasks instructionally related/teachable?
- G. Are the tasks feasible for implementation in your classroom or school in terms of space, equipment, time, cost, and so forth?

5. SETTING THE CRITERIA (RUBRICS)

- A. One or more traits or dimensions that serve as the basis for judging the student response.
- B. Definitions and examples to clarify the meaning of each trait or dimension.

C. A scale of values (or a counting system) on which to rate each dimension.

D. Standards of excellence for specified performance levels accompanied by models or examples of each level.

6. BEGINNING THE DEVELOPMENT PROCESS

A. Investigate how the assessed discipline defines quality performance.

B. Gather sample rubrics for assessing writing, speech, the arts, and so on as models to adapt for your purposes.

C. Gather samples of students' and experts' work that demonstrate the range of performance from ineffective to very effective.

D. Discuss with others the characteristics of these models that distinguish the effective ones from the ineffective ones.

E. Write descriptors for the important characteristics.

F. Gather another sample of students' work.

G. Try out the criteria to see if they help you make accurate judgments about students.

H. Revise your criteria.

I. Try it again until the rubric score captures the "quality" of the work.

Table D 2

Levels of Knowledge

I.	Declarative knowledge (basic knowledge as the foundation)
II.	Procedural knowledge (How to do--process)
III.	Conditional knowledge (When, where, why--justification)

Appendix E Content Expert Agenda

9 - 9:30	<p>Coffee, get acquainted, catch up with each other, share happenings</p> <ul style="list-style-type: none"> -ask Cara to moderate -ask Patty to take notes -sign consent forms -checks
9:30 - 10:30	<p>Present overview of assessment and scoring rubrics Power point presentation</p>
10:30 - 11:00	<p>Present what we did and the process we went through to develop, pilot and evaluate. -Patty and Cara also share their roles</p>
11:00 - 11:30	<p>Have experts do booklet and video assessment? Share student examples and score sheets Validity and reliability How we will rate when we come back Questions and break for lunch</p>
1:00 - 1:30	<p>Review validity and reliability and research question this group is trying to answer</p>
1:30 - 3:00	<p>Evaluation of assessments in station format. Development team is assigned to these assessments to answer any and all questions.</p> <p>1. Station Task (Patty) 2. Booklet (Cara) 3. Video (Researcher)</p>
<p>ORDER of assessment evaluations:</p>	
<p>Reviewers may look at any student examples and all score sheets. Just remember I have promised confidentiality to all parties involved. So please keep student names to self now and through the years.</p>	
John	1. station 2. booklet 3. video
Casey	1. video 2. station 3. booklet
Liz	1. booklet 2. video 3. station

Appendix F
Teacher Instruction
SURGEON GENERAL LESSON CONTENT # 1

1/22/97 9:10 to 9:40
Patty (H5) Fifth

Approximately 10 minutes for this section

1. Talked about permission slips (informed consents)
2. INSTANT ACTIVITY used a Tag game with three taggers. The teacher uses an interesting twist recommended by the children. There are two "cold" taggers that can freeze people and one "hot" tagger that can unfreeze the children. They seem to really enjoy this activity and most students move at a vigorous or medium level of intensity.

Before the activity the teacher DISCUSSED intensity levels with the children.

Teacher: "When we are playing tag and moving by skipping and galloping what speed are you going in, moderate, easy, or hard?"

Student: Moderate

Teacher: "What would be hard?"

Student 1: "Running"

Student 2: "Jogging"

Teacher: "Would jogging be medium or hard?"

Student 3: "Medium"

Teacher shook her head and agreed

Teacher: "What would be easy?"

Student 4: Walking or shuffle step

Teacher: "Right"

3. STRETCHING

Each class period the teacher has a routine of stretching that the youngsters follow. They know the pattern and count to 15 on all the stretches. They do one arm over the head, then the other. Then they do their legs each then both in front. The teacher focuses on their positioning to make sure they do it correctly. Next they do crunches with raised knees. The teacher says up and they raise their shoulders up of the ground not using their head and focus on using their stomach muscles.

4. DISCUSSION: Approximately 10 minutes

Teacher: "I want to ask you some questions before we get started. If someone asked you what does it mean to be healthy for a lifetime what does that mean to you?"

Please raise your hand."

Student 1: "Exercise"

Student 2: "Eat Right"

Teacher: How often do you think it is necessary to exercise, is one time enough to be healthy for a lifetime?"

Coral Response by students: "No"

"How often do you think is necessary?"

Adam?: "Everyday walk 2 miles"

Teacher: "Why would it be okay to walk?"

Student: "Can't do as much if you run. It would equal out if you ran 1 mile each day or walk 2 miles would equal out."

Teacher: "Would moderate or hard be better for you?"

Student: "Moderate"

Teacher: "Why?"

Student 1: "Might strain muscles if you go to hard"

Student 2: "You would get sore and not want to do it anymore."

Student 3: "You would burn a lot of energy"

Teacher: Right, hard or vigorous could be hard on muscles and joints. If you were physically active each day moderate would be appropriate. For example do you think football players practice hard everyday? No, they mix it up and go hard one day, light or moderate in-between days to balance it out."

Teacher: "How much time do you think would be good?"

Student a: 11 to 12 minutes

Student b: 20 minutes

Student c: hour to two hours

Student d: half an hour

Teacher: "You are all in the ballpark. Thirty minutes is appropriate. Do you have to do that thirty minutes all at one time, or can you break it up?"

Student: "Break it up 15 and 15 minutes"

Student: "10 or 5 minutes"

Teacher: 10 minute blocks is okay. Is physically active and physical fit the same thing?"

Student: Coral response: "no"

Teacher: "Right, to be athletically fit or physically fit you would work out harder, but to be healthy it is good to do moderate levels each day for thirty minutes and you can break it up like before, during, and after school."

5. JUMP ROPE STATIONS (approximately 10 Minutes)

Teacher: We will keep working at moderate levels at our stations today.

Station 1. Run through and school 2. Short rope inside long rope 3. Partner jumps
4. Egg Beater with the teacher at this station. There were four children at each group. At the end some students worked on back door and then Double Dutch.

SURGEON GENERAL LESSON CONTENT #2

1/23/97 9:10 to 9:40
 Patty (H5) Fifth

[approximately 10 minutes]

1. INSTANT ACTIVITY & STRETCHING [shoulder, hurdles, legs together and crunches]

2. DISCUSSION

Teacher: "Yesterday we talked about what helps to be physically active and healthy for a lifetime. We talked about thirty minutes a day at moderate level. Today I want to talk about what are some ways you would like to, what are types of things you would choose to do to be physically active if you wanted to be healthy for a lifetime?"

Student: walk or jog

Samantha:

Teacher: "I'll come back to you."

Adam: Weight lifting

Teacher: "If you are weight lifting are you getting cardiovascular health, cause there is a difference between strength and cardiovascular or heart healthy."

Samantha: Nordic track

Boy: Roller blade, mountain bike, soccer, and rock climbing. I have a cliff on my dad's property and I rock climb almost everyday.

Teacher: "Why did you say the things you said. For example why did you say you play soccer?"

Student: I want to do it.

Teacher: "Why did you say rock climbing Albert?"

Albert: Because it's fun.

Teacher: "If somebody told me I had to run 5 miles everyday and I hate to run long distance I wouldn't want to do it. But if someone said let's play volleyball I would do that because I like to play volleyball. So are you going to do stuff you enjoy or will you be more likely to do stuff you don't enjoy?"

Student choral response: If I enjoy it

Teacher: "Are there things you can choose that you would enjoy?"

Albert: "I like to run on the track at Virginia Tech when my dad takes me."

Teacher: "Yes some people like to run. Like there is a guy at my church that loves to run. He says I went 30 miles this week. He really likes to run."

Student: water ski and knee boarding

Teacher: "Yes, skiing and knee boarding can be very fun."

Student: Riding bike

Josh: Street hockey

Rachael: I like swimming and ping pong

Teacher: "What could you do if you can't swim in the wintertime."

Student: I go to the indoor pool.

Teacher: "But, if you couldn't go swimming, what else could you do that you enjoy?"

Student: I go walking and hiking in the winter.

Student: Ice hockey.

Teacher: "What if there is no ice?"

Student: I go to Roanoke to the Civic Center.

Teacher: "I'm sure you can't go there everyday. If you can't go what is something close around here that you do?"

Student: Hunting

Teacher: "When you hunt you walk to get to your spot then you sit for a long time. What other thing do you enjoy?"

Student: Fishing

Teacher: "Physical activity to be healthy for a lifetime does not mean you have to be athletic or have athletic status, but you can still be healthy."

STATIONS

4. [approximately 16 minutes]

Teacher and I worked on back door entrance so students could learn to do double dutch. Many students got into the rope and some were able to jump 2 or 3 times, but most got in and did one jump.

Lesson Plan (Pilot School)

Lesson Plan (Booklet)

1. Instant activity with jump ropes.
2. Levels (Hard, medium, easy).
3. Surgeon General: To be "healthy for a lifetime".
 compared with teacher--leads classroom
 compared with principal--leads school
 compared with president--leads country
 Surgeon General--Leads health of all people
4. Discussion with poster (People doing activities)
 Students discussed what they like
 why?
5. Introduced concepts
 Days
 Intensity why medium? So can continue
 Minutes
 Enjoy--like
 Good at--why?? stick with it
6. Pretend activities
 1 - 2 - 3 Levels of intensity : show me with your body
 Pin points individuals
 Group checking for understanding
7. Closure--Review

Day 2

1. 10 minute activity so students know what it feels like
2. Review with questions
3. Groups partners or 3's
4. Read task--created booklet

Appendix G
Video Tape Advertisement Assessment Task (Weighted Version)

Background

Most children your age do not know how to be healthy through physical activity for a lifetime. A two to three minute video advertisement could help children learn this information in an interesting way.

Task

Your task is to design and create a video advertisement for children which is both very interesting to them and full of correct information.

Audience

The audience for your video advertisement is other fourth and fifth grade students at other schools.

Purpose

The purpose of your work is to use your knowledge of physical activity to entertain and teach children.

Procedure

1. Make a list of the cues you will use.
2. Check the accuracy of your information. Be sure to explain the words.
3. Plan your introduction and conclusion.
4. Make a first draft of the graphics/pictures you will use.
5. Plan a sequence of the activity, graphics/pictures you will use.
6. Practice the order and sequence with your group many times.
7. Practice who will do which parts. Help each other do well.
8. Make revisions and complete the final draft of the graphics/pictures.
9. Make the video advertisement in front of the camera.

**Performance Task Assessment List: Creating a Physical Activity
Video Tape Advertisement for Children**

Student Rubric Score:

- *+ goes beyond what is asked with **quality and accuracy**
- * clearly meets the objective with **quality and accuracy**
- meets the objective but needs improvement
- does not meet the objective

ASSESSMENT POINTS

Content Element	Group Score	Teacher Score	Rater Score
1. The <u>4</u> cues are included in the video advertisement [score by the number of accurate cues included]			
2. Physical activity <u>vocabulary</u> is explained.			
Style Element			
3. <u>Pictures</u> and other <u>graphics</u> support the verbal statements.			
4. The introduction and conclusion catch the interest of the children.			
5. The story line is organized.			
6. The video advertisement is interesting and creative.			

Names of People in your group:

- 1.
- 2.
- 3.
- 4.

Video Tape Advertisement for Children

Teacher & Rater Rubric Score:

4 goes beyond the requirements with **quality and accuracy**

3 meets the requirements with **quality and accuracy**

2 meets the objective but **needs improvement**

1 does not meet the objective

ASSESSMENT POINTS

Content Element	Group Score	Teacher Score	Rater Score
1. The <u>4 cues</u> are included in the video advertisement [score by the number of accurate cues included]			
2. Physical activity <u>vocabulary</u> is explained.			
Style Element			
3. <u>Pictures</u> and other <u>graphics</u> support the verbal statements.			
4. The introduction and conclusion catch the interest of the children.			
5. The story line is organized.			
6. The video advertisement is interesting and creative.			
Content Score x 3			
Style Score			
Total Score			
Total score divided by 9 Average Score			

Names of People in your group:

1.

2.

3.

4.

Appendix H

Student Designed Station Task (Student Directions1)

Background

For this task we want to help you learn what moderate intensity level feels like, so you can maintain that level for at least 10 minutes at a time. A good way to check this is the talk-sing test. If you can still talk to someone but can not sing you are working at the moderate level.

Task

Your task is to design 4 physical activities you can do during physical education class. You will be told what equipment you may use. You will be assigned a space for your group to work in. Choose or design a physical activity that keeps your arms, legs, and large muscles moving continuously the whole time.

Audience

You will rate yourself and one partner during the task.

Purpose

The purpose of your work is to see if you can analyze your own intensity level accurately and correctly. It will also measure whether you can rate your partner's intensity level accurately. Later the teacher and one other person will check to see if their rating matches your rating. Therefore, it is important to try to judge fairly and correctly.

Procedure:

1. Brainstorm ideas for your group.
2. You may select the activities any way your group decides. One idea is for each of you to design one activity all of you can do since your group needs to design a total of 4 activities.
3. Figure out what equipment you will need.
4. Practice doing all 4 activities so you will know you have the equipment and space you will need.
5. Modify the activity or equipment to make it work in the space you have.
6. Use a separate sheet of paper to write or draw a picture of the 4 activities so you will remember them. You will do them during the next physical education class.
7. Towards the end of class time we will teach you how you will record your score and your partner's score.
8. In your next PE class you will all do one station activity at a time. You will rotate to each activity within your group for 1 minute intervals. There will be a 15 second break after each 1 minute interval to record your intensity level and your partner's intensity level.

Student Designed Station Task (Teacher Directions2)

Background

For this task we want to help you learn what moderate intensity level feels like, so you can maintain that level for at least 10 minutes at a time. A good way to check this is the talk-sing test. If you can still talk to someone but can not sing you are working at the moderate level.

Task #1: Your group's task is to design 4 physical activity stations you can do during physical education class. You will be told what equipment is available for you to choose from to be used. You will be assigned a space for your group to work in. Design physical station activities that keep your arms, legs, and large muscles moving continuously the whole time.

Task #2: The next class period, each of you will rate yourself and your partner at the intensity level you worked at doing the station activity.

Audience

Your group will see the finished physical activity station. During the rating part you will be an audience for yourself, your partner, the teacher, and one other adult rater.

Purpose

The purpose of your work is to see if you can analyze your own intensity level accurately and correctly. It will also measure whether you can rate your partner's intensity level accurately. The teacher and one other person will also rate your intensity level. We will check to see if the adult rating matches your rating. Therefore, it is important to try to judge yourself and your partner fairly and correctly.

Procedure:

1. Brainstorm ideas for your group. It is up to your group to decide how to make-up the station activities.
2. You may select the activities any way your group decides. For example, your group may come up with all 4 activities together; or, each of you could design one activity.
3. Figure out what equipment you will need.
4. Practice doing all 4 activities so you will know you have the equipment and space you will need.
5. Modify the activity or equipment to make it work in the space you have.
6. Use a separate sheet of paper to write or draw a picture of the 4 activities so you will remember them. You will do them during the next physical education class.
7. Towards the end of class time we will teach you how you will record your score and your partner's score.
8. In your next PE class your whole group will do each of your 4 activities for 1 minute each. You will all rotate between each physical activity station. There will be a 15 second break after each 1 minute interval to record your intensity level and your partner's intensity level.

Student Checklist for Physical Activity Station Assessment

Name: _____ Date: _____

On each one minute interval the student should code the appropriate intensity level. Your intensity level doesn't matter as much as your ability to rate yourself and your partner accurately and correctly.

M for moderate intensity or higher most of the time

S for moderate intensity or higher some of the time

N for moderate intensity or higher none of the time

SELF ANALYSIS

PARTNER'S NAME:

Minutes	M, S, or N	Minutes	M, S, or N	T	R# 2
1		1			
2		2			
3		3			
4		4			
5		5			
6		6			
7		7			
8		8			
9		9			
10		10			
11		11			
12		12			
13		13			
14		14			
15		15			
16		16			
17		17			
18		18			
19		19			
20		20			

Figure 1.

Student Checklist

Your Name: _____ Date: _____

Partner Name: _____

On each one minute interval the student should code the appropriate intensity level. Your intensity level doesn't matter as much as your ability to rate yourself and your partner accurately and correctly.

H = Hard or vigorous level most of the time

M = Moderate level most of the time

E = Easy level most of the time

	Self Check		Partner	Rater 1	Rater 2
1		1			
2		2			
3		3			
4		4			
5		5			
6		6			
7		7			
8		8			
9		9			
10		10			
11		11			
12		12			
13		13			
14		14			
15		15			

Figure 2. Coding changes from S, M, N (some, most, none of the time) to E, M, H (Easy, Hard, Medium intensity level). See Figure 1 and 2 for details.

Appendix I Stations Teacher Design

Background

For this task we want to help you learn what moderate intensity level feels like, so you can maintain that level for at least 10 minutes at a time. A good way to check this is the talk-sing test. If you can still talk to someone but can not sing you are working at the moderate level.

Task

We will show you the station activities to do. We have picked activities that will keep your arms, legs, and large muscles moving continuously the whole time. You will rate yourself and your partner's intensity level for each one minute interval.

Audience

You will be an audience for yourself, your partner, the teacher, and one other adult rater.

Purpose

The purpose of your work is to see if you can rate your own intensity level correctly. It will also measure whether you can rate your partner's intensity level correctly. The teacher and one other person will also rate your intensity level. We will check to see if the adult rating matches your rating. Therefore, it is important to try to judge yourself and your partner fairly and correctly.

Procedure:

1. We will demonstrate each activity level so you can see the difference.
2. You will listen to the signal so you know when to be active and when to code.
3. We will demonstrate each station for you.
4. You will practice coding students we select to work at different levels.
5. You will get to practice being active and coding yourself and your partner.
6. We will begin and you will rotate to each station after you do your station activity 1 time.
7. At the left of the number code the first letter of the station you are at. For example, if you are at the jump rope station on the fourth (4th) minute, write a **J** to the left of the number 4. We will show you how to do this.

Appendix J Interview Guide Questions for Students

Background & Experience with Assessment

Compared with other children your age, do you think you are more active, about the same, or less active?

What did you think about the alternative assessment task?

Were there parts that were too easy for you? Explain.

Were there parts that were too hard for you? Explain.

Is this similar or different than other tests you do in physical education? Explain.

Likes and Dislikes; Strengths and Weaknesses of the Assessment

Do you think it is important to know this information?

Do you think what you learned about physical activity can help you in life outside of school? How?

Do you think other children your age would like this activity?

What did you like best about the assessment?

What did you least like about the assessment?

Is this information meaningful or important to you?

If you could change the task to make it better, what would you change?

Is there anything, I haven't asked you about this assessment, that you would like to tell me?

Learning

Do you think this task helped you learn anything? Explain

Why did you include the information you included?

How did you remember this information?

Do you think knowing this information will help you know how to stay physically active for a lifetime?

Will it help you stay physically active for a lifetime?

Which assessment made you think the most?

Appendix K
Final Survey

Your Favorite Tasks

Your Name _____ Teacher's Name _____

1. If a friend at another school could only do one of the tasks we did for healthy for a lifetime (Stations, Homer Booklet, or Video Tape Advertisement) which one would you pick for your friend to do?
2. Why?

3. What did you learn from doing your favorite task?

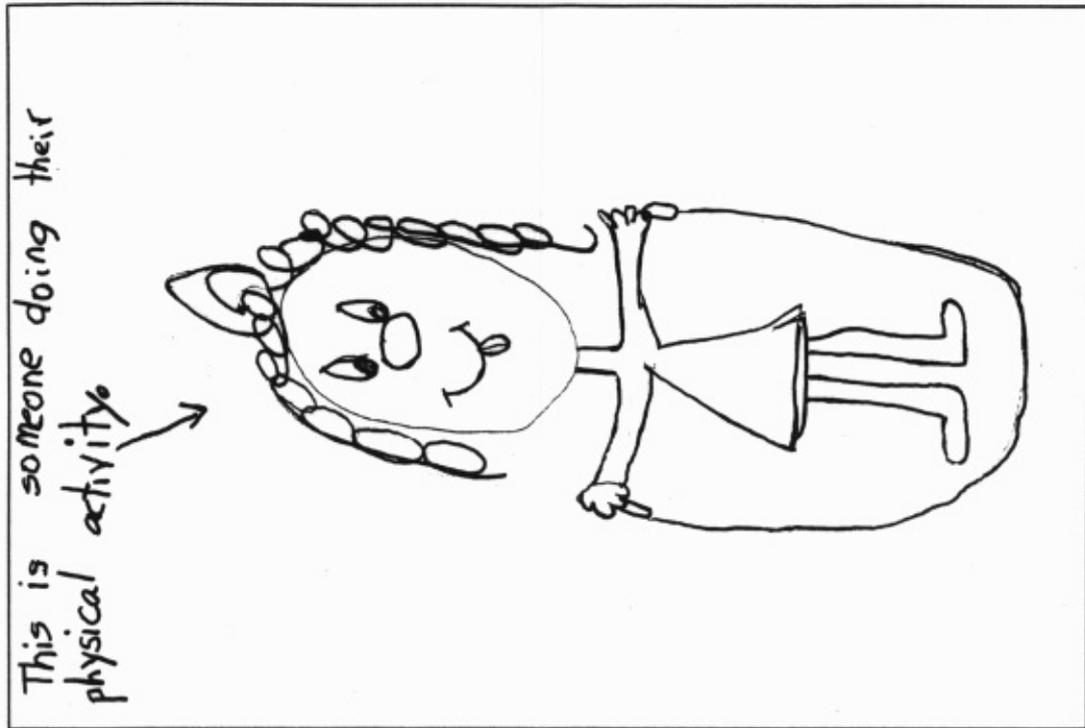
4. Which task best helped you show off what you have learned? Why?

5. Which task challenged you to think the most about how to be physically active for a lifetime? Why?

6. Compared to others your same age and sex, how much physical activity do you get? Circle the answer that best describes you.
 - 5 "much more than others"
 - 4 "somewhat more than others"
 - 3 "about the same as others"
 - 2 "somewhat less than others"
 - 1 "much less than others"

Appendix L

Student Completed Homer Booklet Assessment Task Examples

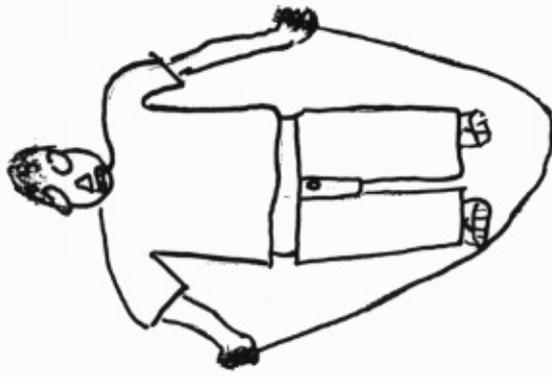


BSS A

Dear Homer,

I heard about your problem so here's the scoop. To be healthy for a lifetime you need to pick something you like for example: basketball or soccer. Now, you need to do that activity 30 minutes a day but you don't have to do it all at once. You could split it into 10 or 15 minutes. You also have to do your activity at a moderate or medium level. What that means is don't do it at a easy level or to hard of level. Also try to do it every day or most days. I hope this helps you!

eat plenty of good foods.
Workout often. Treat your
body well, and don't use
bad things that would
hurt your body.



BHS W

Homes, I heard you didn't know the Surgeon's general warning!
Here's how to be healthy for a lifetime:

- #1. Exercise most or every day.
- #2. Exercise at a moderate level.
- #3. Exercise 30 minutes each day 2 times for 15 minutes or 3 times for 10 minutes or 1 time for 30 minutes.
- #4. Enjoy!

Here are some things you might enjoy:

- 1. Running
- 2. Swimming
- 3. Basketball
- 4. Soccer
- 5. Biking
- 6. Dodgeball
- 7. Kick ball
- 8. Walking
- 9. Skating
- 10. Skateboarding

You list what you'd like to do!

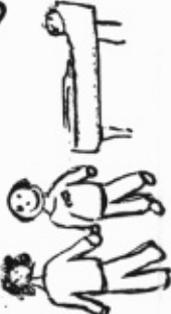
- 1.
- 2.
- 3.
- 4.
- 5.
- 6.

BS4 A

4 Get exercise



To stay healthy for a life time you must

1. Keep your body clean. 
2. Have regular check ups. 
3. Eat healthy. 
- 4 Get alot of sun. 

VITA

I was born in Watertown, South Dakota and attended public schools in Watertown, South Dakota and Bowling Green, Ohio. I also attended American schools in Santos and Sao Paulo, and graduated high school from the American School of Campinas, Brazil.

I graduated from the University of Nebraska-Lincoln in 1982 with a degree in political science. I played golf on the Futures Mini Tour, then pursued a career in business. I returned to school to study elementary physical education in 1986 and graduated from the University of South Florida-Tampa in 1988. I taught elementary physical education for six years in Orlando Florida and worked on my masters degree from the University of Central Florida while I was teaching. In 1994, I took a leave of absence to attend graduate school full time at Virginia Polytechnic Institute and State University in Curriculum and Instruction. I completed my Ph.D. in the summer of 1997.