LOFLIN, JERRY W., Ph.D. Implementation Fidelity of a Physical Education Intervention. (2013)
Directed by Dr. Catherine D. Ennis. 213 pp.

The American educational system has remained relatively stable for the last century as technology, personnel, and evidence-based practices have evolved in schools to serve today's diverse student populations. As researchers continue to design, implement, and evaluate school-based interventions to better serve students, teachers, and schools, complexity theory may serve as a framework for better understanding today's educational system. New school-based curricular interventions are fundamental in meeting the diverse needs of today's students and improving student outcomes. However, the level of intervention effectiveness may be diminished or nullified if a program is implemented in a manner inconsistent with the innovator's plan. Therefore, measuring the extent to which a program implementer faithfully adheres (i.e., fidelity) to the innovator's program ideals during an intervention can contribute to the validation of program outcomes. For this reason, the purpose of this study was to investigate teachers' implementation fidelity and their rationales for changes to the intervention. The findings from this research can assist researchers to effectively anticipate and address factors that may impact fidelity level and student outcomes.

I used a mixed methods design to collect and analyze both quantitative and qualitative data examining implementation fidelity of a school-based physical activity intervention. Data were collected via lesson observations, teacher interviews, student knowledge tests, and accelerometers. Categories and coding themes were developed and quantified from lesson observation field notes to measure teacher fidelity. Multiple

regression was used to analyze the relationship between teacher fidelity and student outcomes (i.e., knowledge growth and physical activity intensity levels). Teacher interviews were coded and categorized to understand teacher rationales for changes to the intended intervention.

A 42-item dichotomous rubric was developed from the open and axial coding of the observation field notes to quantify teacher fidelity scores. Multiple regression with fidelity score as the predictor and knowledge acquisition as the criterion variable, indicated that teachers' fidelity scores accounted for a large portion of variance in student knowledge growth ($R^2 = .79$, adj $R^2 = .74$, $p < .05$). The open and axial coding of the teacher interviews revealed there were a multitude of preexisting contextual factors (e.g., lack of instructional time, space, and equipment) that influenced teachers' fidelity to the intended curriculum. Based on the findings from this research, it appears the more faithful teachers are to teaching research-based curricula as designed, the greater the impact the curricula can have on student achievement. Additionally when designing and revising school-based interventions, researchers should consider the nature of the contextual factors and the extent to which they negatively impact the intervention.

IMPLEMENTATION FIDELITY OF A PHYSICAL EDUCATION INTERVENTION

by

Jerry W. Loflin

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

Greensboro
2013

Approved by

_____
Committee Chair

To my parents, Jerry and Daynese, who have instilled in me the value of hard work and dedication.

APPROVAL PAGE

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

Committee Chair_____
Catherine D. Ennis

Committee Members_____
Ang Chen

_____
Paul G. Davis

_____
John T. Willse

_____
Date of Acceptance by Committee

_____
Date of Final Oral Examination

ACKNOWLEDGMENTS

This dissertation is the result of much support and guidance from my advisor and dissertation committee chair, my dissertation committee, the *Science of Healthful Living* research team, and the administrators, teachers, and students who participated in the implementation of the *Science of Healthful Living* curriculum.

First, I would like to thank Dr. Catherine D. Ennis, who served as my doctoral advisor and dissertation committee chair. Thank you for the constant feedback, encouragement, and support. Your leadership and scholarship in the field of physical education have helped me grow as a student and inexperienced researcher.

Second, I would like to thank Dr. Ang Chen, who served on my dissertation committee. Thank you for your patience and enthusiasm. I appreciate your unique ability to provide constructive criticism with a smile.

Third, I would like to thank Dr. Paul G. Davis, who also served on my dissertation committee. Thank you for your dedication to physical activity, health, and exercise physiology. I appreciate your concern for and interest in school-based physical education.

Fourth, I would like to thank Dr. John T. Willse, who also served on my dissertation committee. Thank you for all the time you spent teaching me about and discussing the world of quantitative research methods. I aspire to have your depth of knowledge.

Finally, this dissertation would have not been possible without the help and support of my peers on the *Science of Healthful Living* research team and the teachers, students, and administrators at the participating middle schools. Thank you to my peers on the research team. I sincerely wish you all the greatest of successes in life. Thank you to the administrators and teachers for allowing me to visit your schools on a regular basis and take time out of your busy schedules. Thank you to the middle school students who enthusiastically completed knowledge tests and wore accelerometers. This dissertation would not have been possible without your participation.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER I

INTRODUCTION

The American educational system has remained relatively stable for the last century as technology, personnel, and evidence-based practices have evolved in schools. Schools have primarily resisted changes in design, curricula, scheduling, and leadership structures (Seltz, 2008). However, the linear, factory-like ideals that historically facilitated school design are proving to be ineffective at meeting current students' needs. "Designed in another time, for the purposes of that time, the typical … school often shows a remarkable lack of flexibility" (Garmston & Wellman, 1995, p. 6). When the modern American educational system was designed in the early 1900s, approximately 51% of 5- to 19-year olds attended school (National Center for Education Statistics, 2012). Today's schools were never intended to meet the moral and economic demands of graduating all students. Furthermore, as the diversity of student populations increases, more poverty-stricken students will be entering America's schools. Tomorrow's students will demand curricula and teaching practices that offset the adverse living conditions that they have experienced since birth (Stevenson, 2010). Today's schools are complex from the classroom level to the district office level and are poorly represented by linear, factory-like frameworks (Davis & Sumara, 2005).

As researchers continue to design, implement, and evaluate school-based interventions to better serve students, teachers, and schools, complexity theory may serve

as a framework for better understanding today's educational system (Fullan, 2008), because it addresses persevering patterns, demonstrated through time and impervious to change. These evolving school patterns can be conceptualized through complexity theory (Fullan, 2003). Complexity theory can function as a theoretical framework for explaining and understanding school-based interventions and recurring themes appearing in schools. When complexity theory is applied to schools, it serves as a metaphor to frame the application of theory within practical contexts. Complexity theory describes multiple levels of interactions that result in opportunities and constraints to the change process within organizations, such as schools (Richardson & Cilliers, 2001). Using such metaphors may be one way for researchers to recognize patterns that impact school-based intervention design.

The Role of School-Based Interventions in School Success

New school-based curricular interventions are fundamental in meeting the diverse needs of today's students and improving student outcomes. The No Child Left Behind Act of 2001 (NCLB; 2003) states that educators must use scientifically based teaching methods and strategies proven to be effective. Proven means that programs have been evaluated through "scientifically based research … that involves the application of rigorous, systematic, and objective procedures to obtain valid knowledge relevant to education activities and programs" (U.S. Department of Education, 2002, p. 7). Interventions provide one method of measuring program effectiveness using rigorous, systematic, and objective procedures. Effective intervention designs use valid and reliable measures of achievement to compare intervention sites that use a particular

program with sites that do not. If the intervention sites produce increased outcomes compared to the sites that do not implement the program, then the intervention/program may be labeled effective. The level of program effectiveness, however, may be diminished or nullified if a program is implemented in a manner inconsistent with the innovator's plan. Therefore, measuring the extent to which a program implementer faithfully adheres (i.e., fidelity) to the innovator's program ideals during an intervention can contribute to program validation. According to Mowbray, Holter, Teague, and Bybee (2003), fidelity is the magnitude to which intervention delivery adheres to the program model originally developed. The measurement of fidelity during efficacy and effectiveness studies is gaining increased interest (National Research Council, 2004).

Interventions are typically designed for implementation in limited contexts with the purpose of increasing programmatic efficacy and internal validity. Once the intervention is disseminated, however, the implementation conditions may be different, or the manner of implementation may vary from the original design. Fidelity measures offer a connection between implementation and outcomes to better understand program effectiveness (Sanchez et al., 2007). Fidelity of implementation and program effectiveness is directly related. Faithfully implemented interventions are likely to produce superior student outcomes compared to those that are not (Burns, Peters, & Noell, 2008; Durlak & DuPre, 2008). In the absence of fidelity measures, intervention designers and implementers may not know if the intervention is effective and if so, in what environments. Most researchers understand that variations in educational interventions occur upon implementation (Dusenbury, Brannigan, Falco, & Hansen,

2003). There is an adjustment point however, that if crossed negatively impacts intervention fidelity. Thus, measuring fidelity, as part of research-based interventions, promotes validity and can explain variance between outcomes. When establishing programmatic validity and interpreting intervention outcomes, researchers who provide data describing fidelity can assist teachers and schools to implement interventions faithfully (Dusenbury et al., 2003).

To maximize fidelity, researchers should identify contextual factors that hinder or facilitate interventions. According to Berman and McLaughlin (1976), three primary variables can impact school-based interventions and implementation fidelity: (a) federal policies, (b) institutional settings, and (c) project characteristics. Federal policies may mandate management theories and objectives, such as the requirement that schools/students meet adequate yearly progress achievement outcomes. Due to adequate yearly progress standards, students may receive more instructional time in some disciplines compared to others, and as a result, impact an intervention's frequency and duration. Institutional settings factors, such as school and district characteristics, organizational climate, and staff motivation, can positively or negatively impact intervention effectiveness. Project characteristics encompassing resource levels, the nature of the intervention, and implementation strategies can increase or decrease intervention outcomes. At the institutional level, teachers are schools' main resource (Wayne & Youngs, 2003) and play a primary role in the faithful implementation of school-based interventions.

Public health and pedagogy scholars have become increasingly concerned about the declining instructional time and overall quality of school-based physical education and physical activity programs. As a result, federal and foundational funding has become available to propose and test school-based physical activity interventions and disseminate effective alternatives to increase physical activity in youth. Research findings describing intervention characteristics and outcomes have been disseminated to inform scholars and practitioners of best practices. Yet, very few school-based physical activity intervention researchers have measured and reported fidelity levels (Sallis et al., 2012). Multiple organizations, such as the U.S. Department of Education, National Institutes of Health, and National Center on Response to Intervention, are requesting implementation fidelity research accompany large-scale intervention proposals. Recently, the National Institutes of Health (2011) established an office of Dissemination and Implementation Research to support future implementation studies.

Statement of the Problem

Fidelity level and intervention effectiveness are directly related to and predict the sustainability of interventions to safeguard positive student learning outcomes (Burns et al., 2008; Durlak & DuPre, 2008). To maximize fidelity, factors that hinder or facilitate the intervention should be identified and measured objectively. Evidence-based research suggests factors related to implementation agents can predict fidelity levels (Berman & McLaughlin, 1976; Brunette et al., 2008; Century, Rudnick, & Freeman, 2010; Datnow & Castellano, 2000; Durlak & DePre, 2008). Teachers are one of the primary implementation agents who impact implementation effectiveness through their conscious

and unconscious decisions to implement the curriculum with various levels of fidelity. For this reason, the problem addressed in this study was to investigate teachers' implementation fidelity and their rationales for changes to the intervention. The findings from this research can assist researchers to effectively anticipate and address factors that may impact fidelity level and student outcomes.

Purpose of the Study

Measuring fidelity and examining its impact on student outcomes is a promising concept in educational research (Mowbray et al., 2003; O'Donnell, 2008). The purpose of this study was to examine sixth-grade physical education teachers' fidelity levels when implementing the *Cardio Fitness Club* unit from the *Science of Healthful Living* curriculum and to investigate the relationship between teacher fidelity levels and student outcomes (i.e., knowledge growth and physical activity intensity levels). I also investigated the extent to which teachers adapted or ignored critical intervention components and design principles and teachers' perceptions and justifications for these changes as they influence fidelity level.

Research Questions

This research was guided by three research questions:

1. To what extent do teachers implement lessons from a sixth-grade physical education curricular intervention with fidelity (adherence – coverage and delivery quality)?
2. What is the relationship between implementation fidelity and student outcomes (i.e., knowledge gain and physical activity intensity levels)?

3. What reasons do teachers offer for making changes to the intended curriculum as they implement it in their school contexts (moderating variable of context)?

Significance of the Study

Assessing implementation fidelity increases researchers' and practitioners' understanding of the fit between the intervention and the context. However, the complexity of the implementation context can impact the efficiency and generalizability of the innovation to other situations. Researchers have rarely used fidelity variables to explain intervention outcomes. Appropriate and accurate fidelity data can inform research and practice to better explain intervention outcomes. Further, when researchers understand teacher justifications for adapting intervention components, they can design or modify interventions to accommodate diverse contexts, minimizing factors that negatively impact implementation.

During a time of teacher and student accountability, high stakes testing, and No Child Left Behind mandates, schools and students must have effective curricula to increase achievement outcomes. Furthermore, given the calls by public health organizations to increase physical activity and decrease overweight and obesity levels in children and adolescents, school-based physical education programs need evidence-based interventions proven to increase educational achievement and /or physical activity outcomes. Measuring implementation fidelity provides one method of assessing moderators of curricular effectiveness. According to O'Donnell (2008) there is need to conduct research investigating implementation fidelity and to understand why and how teachers impact intervention implementation. Studying implementation fidelity as part of

curricular effectiveness research is critical to understanding intervention success or failure. Valid measures of implementation fidelity also can improve statistical power and explain variance in outcome studies. The data gathered in this study provided information about the variables that enhanced or constrained the effectiveness of a large-scale physical education intervention, the influence of implementation fidelity on student outcomes, and teacher rationales for intervention adaptations within their context.

Assumptions

I assumed that the integration of qualitative and quantitative research methods within the same study complemented each other, generating supportive evidence useful when drawing conclusions and making informed decisions. Further, I believed that physical education would continue to be a middle school subject. The existence of physical education in schools has been jeopardized in the past, but has managed to survive. Additionally, this research was based on my assumption that teachers would participate (i.e., planning lessons, teaching lessons, and completing surveys) in the study with integrity and to the best of their abilities.

Limitations

This dissertation study should be viewed with some limitations. First, the participants (i.e., the teachers) in the study were not randomly selected although schools were randomly assigned to experiment and control conditions. Second, the teachers in this study participated for the first time in the intervention. I observed only 3 of the 15 schools in the experimental condition. Additionally, within these three schools I observed only 30% of the lessons per sixth-grade class. Fidelity data from 30% of the

lessons and 20% of the experimental schools may not provide a representative assessment of teacher fidelity levels and instructional practices across the entire intervention. Further, students were not selected randomly to wear accelerometers. Instead, they were selected by teachers to represent specific student characteristics and demographics. Thus, physical activity intensity data from six students per sixth-grade class per school (n = 102 students) may not provide a comprehensive description of average student physical activity intensity levels in the intervention. Finally, student knowledge growth data were limited by the reliability and validity of the knowledge tests. These concerns may limit the utility and generalizability of the study findings.

Delimitations

Several delimitations exist in the study. First, due to the large number of potential participants in the study population, participation in the study was delimited to teachers who taught sixth-grade physical education in the Piedmont region of North Carolina and taught one unit of the two-unit intervention for the first time. Second, the study was delimited to the examination of teachers' fidelity levels when implementing a specific science-enriched physical education curriculum, the *Science of Healthful Living*. Third, implementation fidelity was measured using diverse measures, including qualitative coding and themes from lesson observations.

Definitions of Key Terms

To ensure clarity and understanding throughout the study, I define the relevant key terms in the following paragraphs.

Adaptations

Adaptations are changes made in the original program during implementations that include both modifications and omissions (Durlak & DuPre, 2008).

Adherence

Adherence refers to the extent at which an agent delivers a unit as designed or written (e.g., implementation of activities and methods) (Lynch, 2007; Lynch & O'Donnell, 2005).

Attractor States

Within complexity theory, attractor states in a complex system are patterns that emerge between variables. Complex systems are either maintaining or flowing to an attractor state (Handford, Davids, Bennett, & Button, 1997).

Bifurcation

Within complexity theory, bifurcations are small, critical unstable points within a complex adaptive system that move a complex adaptive system from one attractor state to another (Capra, 2002).

Chaos Theory

Chaos theory is the qualitative study of unstable aperiodic behavior in deterministic nonlinear dynamical systems. Systems are dynamic (i.e., changes over time), aperiodic and unstable (i.e., does not repeat itself), can have simple causes, nonlinear and sensitive to initial conditions, and are neither random nor predictable (Kellert, 1993). Complexity theory emerged from conceptualizations of chaos theory.

Complex Adaptive Systems

Within complexity theory, complex adaptive systems are non-linear, complex, interactive systems that have the capability to adjust to a fluctuating environment. They are characterized by continuous and discontinuous change, emergent phenomena, and unpredictable outcomes. Examples include the nervous and immune systems, living organisms, and social systems (Goldstein, 2008).

Complexity

Complexity is a flow of activation states between extremes of variation and sameness, chaos and order. It is attained within a system's fluctuating conditions by combining variation with integration. Systems moving toward complexity are most stable, flexible, and capable of a wide range of self-organizing processes (Siegel, 2001).

Complexity Science

Complexity science expands systems thinking derived from organismic biology to non-living systems. Complexity scientists assume that complex systems have some behavioral and structural features in common and can be represented mathematically (Capra, 1996; Thelen & Smith, 1994).

Complexity Theory

A theory describing phenomena demonstrated in systems characterized by nonlinear interactive components, emergent phenomena, continuous and discontinuous change, and unpredictable outcomes (Clancy, 2004). Complexity scientists examine complex systems using complexity theory.

Context

Context describes the surrounding social structures, such as customs and cultures of organizations and groups, inter-organizational linkages, and historical as well as concurrent events, in the intervention setting (Lipsey & Cordray, 2000). Context may refer to the setting in the previous or former program, described in this research as the preexisting context or the context created by the new intervention, referred to as the intervention context.

Coverage

Coverage denotes delivery completeness or the content amount received by participants (Dusenbury et al., 2003). Also referred to as exposure, dose, or dosage, this dimension may include the number of sessions, length of each session, or frequency of implementation (Dane & Schneider, 1998).

Critical/Core Components

Critical or core components are the essential elements of the explicit and/or implicit program model, such as elements that constitute a particular intervention, innovation, or program, without which the integrity or effectiveness would be compromised (Century et al., 2010).

Delivery Quality

Delivery quality assesses the extent to which the provider approaches the developer's idealized theoretical framework in content deliverance terms (Dusenbury et al., 2003). Lynch and O'Donnell (2005) describe it as how well the teacher implements

the unit: evidence of theoretical ideals, skills, enthusiasm, preparedness, and delivery attitude in a manner consistent with the developer's intentions.

Effectiveness

Effectiveness is defined as the ability of an intervention to produce the desired beneficial effect in actual use (Dorland, 1994, p. 531).

Efficacy

Efficacy is the ability of an intervention to produce the desired beneficial effect in expert hands and under ideal circumstances (Dorland, 1994, p. 531).

Emergence

Emergence describes the patterns and properties that arise from the spontaneous organization of multiple subsets through their interactions, sometimes leading to novelty in a system (Axelrod & Cohen, 2000).

Fidelity of Implementation (FOI)

Fidelity of implementation is the extent to which the delivery of an intervention adheres to the program model and theory as originally developed. It includes content, frequency, duration, delivery quality, and coverage (Lynch, 2007; Mowbray et al., 2003; Ruiz-Primo, 2006).

High Implementation Fidelity

High implementation fidelity refers to the implementation of the essential intervention components (Carroll et al., 2007).

Intervention

An intervention is the treatment in an experimental or quasi-experimental trial designed to change program outcomes. Interventions introduce new conditions to a preexisting context or program.

Outcome Analysis

Outcome analysis is a process by which the result(s) of an intervention are formally assessed.

Participant Responsiveness to Intervention

Participant responsiveness is viewed as the participants' levels of engagement in the intervention (Dusenbury et al., 2003). Responsiveness also refers to the degree to which the program stimulates the interest or holds participants' attention (Durlak & DuPre, 2008).

Self-Organization

Within complexity or chaos theory, self-organization is a process through which some form of global order or coordination arises out of local interactions between components of an initially disordered system.

Strategies to Facilitate Implementation

Strategies to facilitate implementation are support strategies used both to optimize and to standardize implementation that may include the provision of training, feedback, manuals, guidelines, and monitoring provided to intervention implementers by developers for the purpose of enhancing fidelity of implementation (Carroll et al., 2007).

CHAPTER II

REVIEW OF THE LITERATURE

The purpose of this review is to analyze the complex web of variables that interact in the faithful delivery of complex school-based physical activity interventions. Because schools are socially and academically complex entities with numerous layers of authority and resistance (Morrison, 2006), a primary challenge when examining questions of intervention faithfulness, or fidelity, is the identification and modification of relevant contextual phenomena that significantly impact student outcomes. One theory that provides insight into sophisticated dynamic systems, such as schools, is complexity theory (Morrison, 2006). Utilized effectively by researchers in academic disciplines, such as mathematics and physics, to examine systemic change (McMillan, 2008; Stacey, 2003), complexity theory was employed in this research as a metaphor to better understand systemic complexity as it applies both to stable school contexts and to the large-scale curricular interventions that attempt to change them.

Within curricular study of large-scale interventions, implementation fidelity poses a vexing problem. Fidelity, or the extent to which an intervention is implemented faithfully, is a construct that is sensitive to political, institutional, and pedagogical factors that both facilitate and inhibit change (Durlak & Dupre, 2008). An improved understanding of curricular fidelity is essential if researchers are to increase the outcomes and sustainability of large-scale school interventions leading to educational reform.

Therefore, this chapter is divided into four sections: complexity theory, fidelity implementation and measurement, institutional barriers and opportunities for school reform, and school-based physical activity interventions.

Complexity Theory

Traditionally, scientists have viewed the world through a lens of reductionism. In other words, reductionism involves breaking down whole entities into their subsequent parts. Historically, reductionists have hypothesized that if scientists understand the nature and function of individual system components; they will understand the workings of the whole system (Coveney & Highfield, 1995). Recently, some researchers have moved beyond reductionism to examine the whole entity to determine how it functions without reducing it into smaller pieces. Chaos theorists, for example, consider the whole entity without reduction and propose that an iteration of a small change within the large system can result in very complicated outcomes (Elliot & Kiel, 1997). Complexity theorists take this principle one step further considering the whole system, but suggest that complex systems can generate rich, dynamic behavior from simple interactions between many subunits. Schools are an example of a complex system containing many interacting subunits (e.g., students, school personnel, policy mandates, and institutional context). The interacting subunits determine unique school characteristics that cannot be reduced to the individual subunit characteristics, alone (Fullan, 2008). Thus, faithfully implementing a new curriculum depends on more than the subunits and their interactions. It also depends on the unique characteristics of the school as a whole. Complexity theory

provides a framework to better understand and explain the faithful implementation of interventions in schools.

This following section outlines the utility and potential of complexity theory and discusses applications, developments, and features, as well as some major contributors to the field. To better understand complexity theory, concepts from chaos theory will be used as a foundation facilitating an understanding of the section focus, complexity theory.

Chaotic Systems

A chaotic system is similar to a weather system. It displays patterns and stability on the large scale, while its less noticeable behaviors are never repeated in episodic forms. Chaotic systems are stable and their actions can be perceived as real occurrences. They display, however, dynamic actions because they are sensitive to initial differences in system conditions (Coveney & Highfield, 1995; Elliot & Kiel, 1997). The core of chaos theory explains that system performance is not founded on elementary cause-and-effect interactions where results can be anticipated as a function of contributions (Marion, 1999). Instead, a distinct action taken to influence a system understood within chaos theory may generate any of three alternatives: no reaction at all, an equal and opposite reaction, or an unequal and unbalanced reaction (Gleick, 1987). Chaos theory contradicts traditional Newtonian motion physics, arguing that strange attractors can facilitate unpredictable interactions among system units in chaotic systems (Gleick, 1987).

The most frequently cited chaos theory example is the alleged "butterfly effect," initially offered by meteorologist Edward Lorenz at the Massachusetts Institute of Technology in the early 1960s. In this metaphor, the movement of a butterfly's wings on

one side of earth is theorized as having an unequal, advancing effect on the weather system macro-dynamics on the opposite side of earth (Gleick, 1987). As suggested by the butterfly effect, open systems in nature maintain a delicate dependence on initial conditions. Therefore an insignificant input can echo throughout a natural system, causing a result that varies considerably from one result that would have ensued without the input (Elliot & Kiel, 1997).

Chaos theory has evolved informing more recent conceptualizations of dynamic systems known as complexity theory. Currently complexity theory is being employed by pedagogical scholars (e.g., Davis and Sumara, 2005) to better understand the complex dynamics of schools. Bearing in mind,

> Post-normal science does not replace good quality traditional science technology. It reiterates, or feedbacks, their products in an integrating social process. In this way, the scientific system will become a useful input to novel forms of policy-making and governance. (Sardar & Abrams, 2008, p. 159)

Thus, a framework understanding is critical to capturing the capability to utilize complexity in meaningful ways.

Complexity Theory

Complexity theory demonstrates how trivial behavioral patterns can evolve from numerous exceedingly complicated and dynamic relationships. Like chaotic systems, complex systems also have distinctive characteristics, such as the capacity to balance between order and chaos or be on "the edge of chaos" (Langston, 1990). Complex systems function between instability and stability (Sardar & Abrams, 2008). Complexity theorists postulate that large, dynamic systems possess three common characteristics:

self-organization, emergence, and adaptation. Systems in balance demonstrate emergence and self-organization. Self-organization is the system's ability to re-define itself without input from outside sources. Emergence describes the properties and patterns that evolve from multiple subsets organizing through their interactions. Adaptation occurs when a system adapts through a selection process, facilitating improvement in one or more success measures (Axelrod & Cohen, 2000).

Complex adaptive systems. Complex adaptive systems are dynamic, interactive entities that bifurcate and stabilize in unpredictable ways. They reject Newtonian science while taking a responsive and interactive stance. The observer's participatory role in the system fosters subjectivity. Complexity definitions contradict regularity, predictability, and order, and what it means to live in time (Peat, 2002).

Thus, complex adaptive systems are systems containing agents with the ability to interact and change (Axelrod & Cohen, 2000). For example, teachers, students, and administrators are the primary agents in schools. Such a multi-agent system contains agents that interact locally but produce a higher-order global pattern. "Agents are intrinsically subjective and uncertain about the [global] consequences of their actions, yet they generally manage to self-organise into an emergent, adaptive system" (Heylighen, Cilliers, & Gershenson, 2007, p. 132).

Initial conditions. The evolution of a system's progress depends heavily on its initial conditions. As mentioned previously, Edward Lorentz first recognized the powerful role of initial conditions while investigating weather models (Gleick, 1987). While using a computer simulation program to better understand weather patterns,

Lorentz discovered that starting a weather simulation at a point other than the beginning produced unique results each time the simulation was run from that same point (Capra, 2002). Thus, prediction within the simulated weather system was not feasible. As a result, Lorenz postulated that making small changes to initial conditions could lead to large changes in the way patterns develop.

Attractor states. In complex adaptive systems agents follow a few simple rules. Therefore, patterns evolve over time and become *attractor states* in the complex systems. Complex systems either move toward attractor states or maintain an attractor state (Handford et al., 1997). Given the sensitivity of a complex system, small changes in one system agent may create large, dynamic changes in the system's behavior. These patterns are difficult to predict, but they do not happen by chance. As small system changes create large-scale, dynamic behavior, discontinuities emerge. These discontinuities are called *bifurcations*.

Bifurcation. Bifurcation points are places of turbulence in a complex system (Capra, 2002). As attractor states change, emerge, or disappear, system stability decreases. Small changes may disturb a complex adaptive system in such a way that a bifurcation emerges as a new attractor state. New patterns and stability evolve at these bifurcation points (Capra, 2002).

Complexity theory offers a novel, useful theoretical framework from which to examine the adaptations of large-scale, school-based interventions. Understanding complex system characteristics permits their application as metaphors in education to conceptualize intervention challenges from a different perspective. These metaphors may

offer unique, progressive ways of thinking about schools and school-based interventions. As schools continue to be challenged to meet outcome-based standards and accountability mandates, having new metaphors may help address and promote positive changes in new ways.

Complexity as a Metaphor

Complexity science offers a unique perspective on schools and may allow stakeholders of large-scale, school-based interventions to better understand changes in schools. Complexity theory can serve as an influential theoretical model, a metaphorical tool, or a powerful tool through which to examine organizations, such as schools (Richardson & Cilliers, 2001). It interprets complex organizational phenomena in a unique, non-traditional way, concentrating on the complete organization (Kauffman, 1995).

The classroom and school setting is one example of a complex system. As an often disregarded complex adaptive system, this system consists of students, teachers, staff, and administrators merging "… into a unit of cognition whose capacities exceed those of the individuals on their own" (Davis & Sumara, 2005, p. 315). Schools represent a multitude of interactive sub-units working as a whole unit and serves as a good example of a complex adaptive system.

Since schools are primarily made up of people, human organizations also can be investigated via a complexity theory lens examining attractor states and initial conditions. If human organizations, such as schools, are well-defined as complex adaptive systems,

then minor changes in schools may facilitate a dynamic change from one attractor state to another.

Initial Conditions and Attractor States in Schools

The American school ideal has primarily remained the same over the last century and resisted change. The initial conditions in schools during Horace Mann's and John Dewey's times may still be seen in today's schools as stable educational trademarks. According to Meyer and Rowan (1983), the American educational system represents history and is resistant to change. Examples include the taken-for-granted structures in schools such as the organization and timing of the school day, grade level organization, teacher certification, school appearance, and even which subject areas will be taught and tested. Assuming that these particular combinations create a "school" leads to educational structures that are "nonconfrontable and nondebateable, and hence are extremely difficult to change" (Schein, 2010, p. 28). Historical school characteristics such as school building design, grade level assignment, school day scheduling, and instructional design, have operated as initial conditions for the American educational system. An awareness of these conditions is essential if teachers, administrators, curriculum specialists, and researchers are to work within a system experiencing bifurcation points due to, for example, the changing student population or the implementation of a new curriculum. The success of large scale, school-based interventions depends not only on their ability to change the whole, but also on their ability to positively change initial conditions and attractor states.

Similar to weather, the initial conditions historically present in the American educational system have molded schools into their present state. These initial conditions are stable, and continue to underpin educational systems today. Some initial conditions (e.g., scheduling and instructional design) are so stable that new interventions or curricula must be positioned carefully to fit within or purposely cause bifurcations in these exceptionally stable systems.

Complexity in Physical Education

Similar to schools and education in general, physical education as a school discipline historically has displayed many of the same stable characteristics resistant to change. However, there is only a small body of literature articulating complexity theory in physical education teaching, curriculum, and research. Nevertheless, complexity theory offers many benefits to structuring and explaining physical education practices that can lead to successful implementation of complex, large-scale interventions.

Ennis (1992) was one of the first scholars to analyze physical education curriculum design and learning through a complexity lens. The notion that learning in physical education could occur on "the edge of chaos" contradicted prevailing physical education curriculum specialists' beliefs that physical education curriculum design and learning occurred in linear, stable environments. Ennis posited that multiple variables (e.g., attractors) in physical education settings influence learning, facilitating or hindering the learning process from reaching a state of homeostasis. This "far-from-equilibrium" physical education learning environment creates challenges for physical education teachers as they attempt to identify the major attractors affecting their teaching and

students' learning. Identifying these challenges related to physical education teaching and learning can allow physical education curriculum developers to target interventions to bifurcate the learning process increasing student outcomes (Light, 2008).

Over the last 10 years, several scholars have proposed physical education curriculum models, such as Teaching Games for Understanding, sport education, and movement education, informed by complexity theory (Light, 2008; Light & Fawns, 2003; Storey & Butler, 2012). These models not only provide students opportunities to move, but also opportunities to think about their movement. The integration of the brain and human body movement represent self-organizing, complex adaptive systems (Morrison, 2008). Learning in physical education involves understanding the body's responses (i.e., physiological changes) to physical activity, why and how (i.e., biomechanics) the body needs to move in particular pathways for different tasks, and the complex relationships within and between the two (Light, 2008). Physical education classes, like larger organizations, have initial conditions and attractor states that can allow classes to appear linear and stable. Teaching Games for Understanding, sport education, and movement education curriculum models provide only general rules for movement, so students are free to improvise widely. Similar to a bird flock, the general class characteristics can be anticipated, but individual class experiences will be diverse. Such responses are not created or directed by individuals, but are emergent within the system.

Physical education is a collective effort representing the complex interactions within and across educational systems' components (Jess, Atencio, & Thorburn, 2011). The dynamic, unstable components, such as teachers' and students' beliefs and

personalities and preferences, can create engaging learning opportunities that produce new knowledge for students (Osberg, Biesta, & Cilliers, 2008). In one example, Jess et al. (2011) developed and delivered a physical education program in Scotland between 1999 and 2006. Based on observations and teacher feedback, they began restructuring their physical education program in 2006 to demonstrate the feasibility of complexity theory as a framework for learning in physical education. The new model focused on students' general, as opposed to sport specific, movement patterns and the important social, cognitive, and developmental factors that impacted learning in physical education. The new model was introduced over a prolonged time period to gradually replace the traditional multi-activity approach to teaching. Additionally, the new model provided students opportunities to actively participate in self-organizing practice. As described, the development and implementation process for the new curriculum was complex due to resistant teachers and traditional physical education practices. To implement the changes, the physical education teachers were required to change from the perspective of "one correct movement form," to an emergent, non-linear exploration of "best fit" movement forms. Student instructional tasks were changed to encourage more independent, self-organizing, student-centered practice to enhance student learning and move away from stable, teacher-centered lessons (Jess et al., 2011). The restructured developmental physical education program outcomes and results are unknown at this time, as the intervention was still in its implementation phase at publication. Nevertheless, Jess et al. (2011) do provide a unique curriculum restructuring example

using complexity theory as the framework that can be used as a guide for restructuring additional teacher-centered curriculum models.

Most complex systems fluctuate between stability and change. Rarely, however, do *entire* systems change. Instead, change occurs in different elements of the system at different times. Bifurcating subsystems evolve into new initial conditions and can produce new routines, beliefs, and languages that individual agents must learn and embrace if the new system is to again reach temporary homeostasis.

Historically, schools and school districts have attempted to implement thousands of new programs. Before a new curriculum can become stable and predictable, however, agents responsible for implementing the change need to understand and embrace the new programs' beliefs, strategies, and technologies. For researchers to examine the impact of the implementation, agents must first accept the program as presented by the developers. The first step in the examination of evidence-based practice is to collect data to support the supposition that the new standardized program is being implemented faithfully as designed. However, ensuring implementation faithfulness, or fidelity, within diverse, complex environments is one of researchers' greatest challenges. Implementation fidelity should be measured to confirm that the outcomes of the new curriculum are attributable to that curriculum operating as designed with the complex system and not to other unidentified, ill-defined, or unmeasured variables.

Fidelity of Implementation

Rogers (2003) pointed out that until the 1970s researchers found no need to study fidelity of implementation (FOI). They assumed that those who implemented a new

program would do so with fidelity and implementation would mirror previous adoptions. As it became more obvious that program implementers regularly altered a program or intervention to meet individual and institutional barriers and opportunities, researchers began to conduct research to examine the context surrounding the fidelity construct. The agents that compose a complex adaptive system are the same agents that impact a program's FOI when that program is introduced to the complex adaptive system. The initial conditions present in both the complex adaptive system and the program, itself, have the ability to interact within and between agents producing patterns or attractor states. Since complex adaptive systems are sensitive to change, any change to or by the program may modify the personality of the system and negatively impact FOI.

Although FOI has been widely examined in the health field, fidelity has not been the program evaluation focus in most educational studies (O'Donnell, 2008; Ruiz-Primo, 2006). Nevertheless, according to the U.S. Department of Education Institute of Education Sciences' (2011) position in the *What Works Clearinghouse* report, intervention effectiveness depends mainly on adherence to the program's guiding principles. For example, the moderating variable "reduced class size" is a program requirement that, when carried out faithfully, has been shown to be effective in improving academic achievement outcomes. However, when the program requires reduced class sizes and the institution continues to implement programs with typical large class enrollments (low fidelity), the program efficacy declines significantly.

Fidelity Components and Moderators

Arriving at a consensus on a fidelity definition is central to this evolving area of research. Century et al. (2010) completed a comprehensive analysis of fidelity definitions, concluding that fidelity could be operationally defined as "the extent to which the critical components of an intended program are present when that program is enacted" (p. 207). Further, Carroll et al. (2007) review of implementation fidelity research identified six components: adherence, coverage, delivery quality, participant responsiveness, context, and strategies to facilitate implementation. A model of the fidelity construct is presented in Figure 2.1. The dashed line in Figure 2.1 signifies that the relationship between an intervention and its outcomes is external to implementation fidelity, but the level of implementation fidelity attained can influence this relationship. Carroll et al. (2007) further defined these components within the fidelity context. Specifically, *adherence* refers to the degree to which implementers are delivering a program as it was intended. *Coverage* denotes delivery completeness or the content amount received by participants compared to what was prescribed, while *delivery quality* refers to how well agents are implementing intended techniques and methods in the program. *Participant responsiveness* is the extent to which the participants are engaged in the implemented program. *Context* describes the surrounding social structures, such as customs and cultures of groups and organizations, inter-organizational linkages, and past as well as simultaneous events, in the intervention setting. Finally, *strategies to facilitate implementation* are support strategies used both to enhance and to standardize implementation that may include the provision of training, feedback, manuals, guidelines,

and monitoring provided to program implementers by developers for the purpose of enhancing FOI (Carroll et al., 2007). These six components are discussed in greater detail later in the chapter.

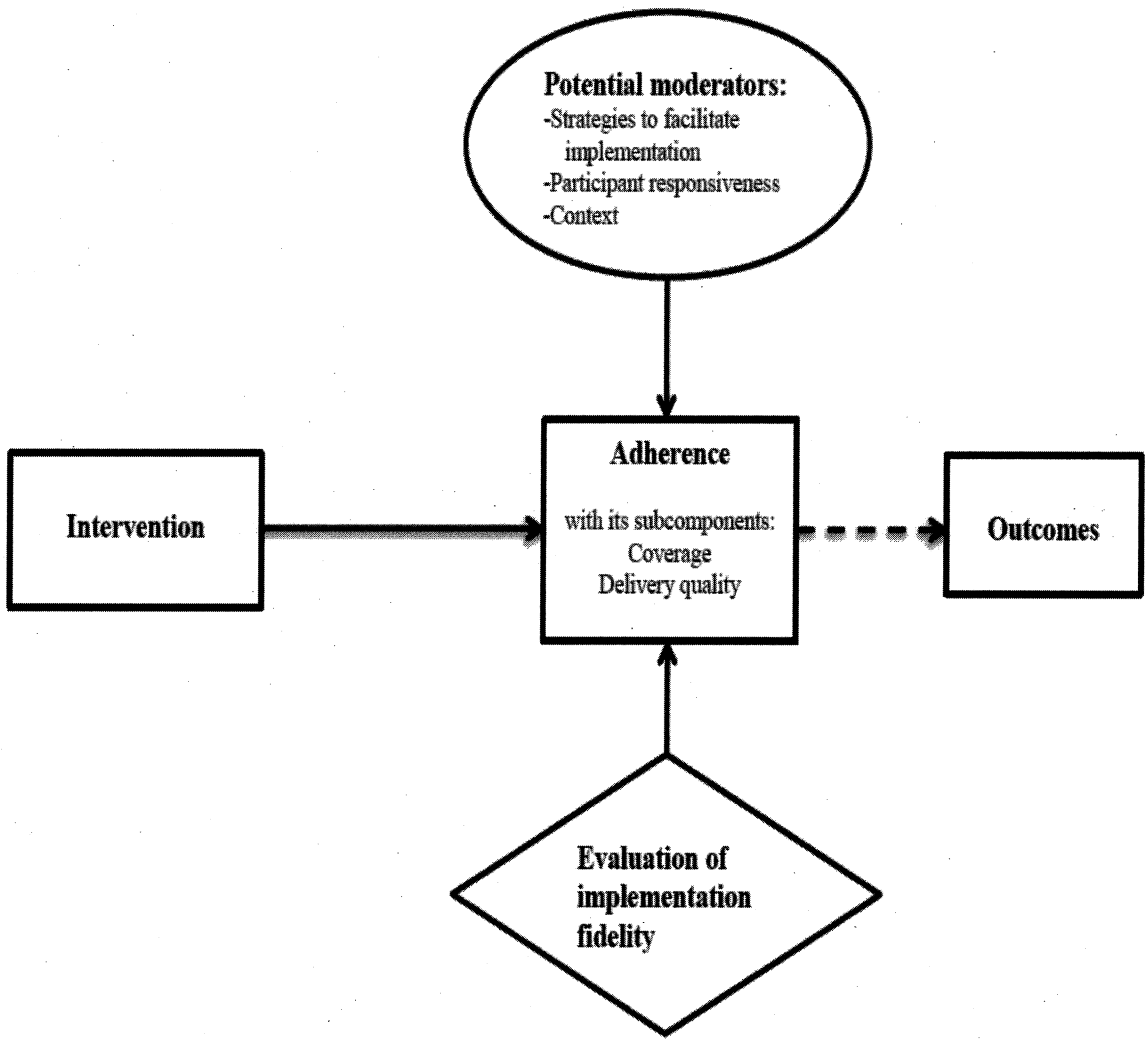

Figure 2.1. Modified conceptual framework for implementation fidelity (Carroll et al., 2007).

Dane and Schneider (1998) emphasized that fidelity consists of all six components and comprehensive studies of fidelity should measure all six to ensure a

complete program or intervention fidelity construct. Other researchers, however, target particular concepts to include based on individual study's needs and requirements. Century et al. (2010) further suggested that the components and subcomponents of the *adherence* component can represent FOI with the other three concepts providing explanations and degrees of enthusiasm, interests, and participation.

In addition, Ruiz-Primo (2006) identified some intervention factors that may limit implementation fidelity. For example, *intervention complexity* includes three elements: the total number of intervention components that require implementation, the intervention's similarity to the previous program, and teachers' prior knowledge of intervention components. The *amount of time* necessary to implement the intervention also is a factor, with longer duration correlated to less fidelity. In other words, interventions that require more time to implement may be challenging to implement with fidelity. Conversely, a small number of required resources critical to the intervention's success, as well as the utility of the supplied implementation manuals, are both linked to higher fidelity. *Training and participant buy-in* to the intervention also are essential components. Finally, Ruiz-Primo (2006) stated that *intervention provider supervision* is a crucial factor in FOI, particularly when applicable feedback is provided regularly.

Managing these factors to promote high fidelity and intervention success is a challenging task. Current research provides very little data concerning the ideal intervention length, number of components, or degree of similarity to the previous program. Increasing learning outcomes require instructional time, resources, and administrative support, and any quality educational intervention with these goals requires

the same. Typically schools are stable organizations, and as such, an intervention needs to address school and programmatic complexity effectively to overcome traditional school stability. Faithfully implemented interventions also require a balance of the factors mentioned above. For example, interventions need to be long enough to enhance learning, but not so long to increase teacher and student burnout. Duration, however, is unique to the preexisting school context. Interventions need to be sufficiently unique to be unlike the previous program, but similar enough to the previous program that teachers understand, embrace, and facilitate the intervention with confidence and success.

Fidelity of Implementation Efficacy and Effectiveness

In outcomes-based educational research, FOI instruments are used to examine the extent to which the results can be attributed to the intervention. For many educational interventions, the purpose is to improve student achievement outcomes. FOI instruments are administered during efficacy and effectiveness portions of a study to ensure FOI at different stages (Mills & Ragan, 2000) and to account for the differing study foci (O'Donnell, 2008). *Efficacy* can be defined as "the ability of an intervention to produce the desired beneficial effect in expert hands and under *ideal* circumstances" (Dorland, 1994, p. 531). Thus, efficacy studies focus on the degree to which an intervention is implemented and successful in an expert-created ideal environment. This type of formative assessment information is used to make improvements to the intervention. Therefore, during an efficacy study, fidelity is a component of internal validity. The process involves critical components identification (Mowbray et al., 2003), continuous

monitoring (Resnick et al., 2005), and programmatic revisions as necessary during each phase (Resnick et al., 2005).

Conversely, *effectiveness* is defined as "the ability of an intervention to produce the desired beneficial effect in *actual* use" (Dorland, 1994, p. 531). During an effectiveness study, FOI is measured to determine external validity or generalizability of the intervention. Effectiveness measures investigate the intervention's ability to produce the desired outcome in real-world situations (O'Donnell, 2008). Without FOI studies, researchers cannot determine the extent to which outcomes are attributed to implementation of the intervention framework (Dobson & Cook, 1980). According to Greenberg et al. (2005) and O'Donnell (2008) effectiveness studies can determine the capacity for scale-up (i.e., expanding efficacious interventions from small settings to multiple settings).

Fidelity of Implementation Monitoring Stages

FOI can be examined and monitored during three stages of outcome research: (a) pre-adoption phase, (b) delivery phase, and (c) post-delivery phase (Greenberg et al., 2005). During each phase, researchers investigate distinct issues that should be addressed to facilitate increased FOI. Specifically, during the pre-adoption phase, intervention designers evaluate the curricular fidelity with the intervention's theory base. To promote higher fidelity levels, the intervention designers (a) examine program fit with participants' needs, goals, philosophy, and organization; (b) secure sufficient resources to conduct the intervention as intended; (c) train implementers and observers; and (d) set up

a supportive, problem solving atmosphere to facilitate trust and communication (Greenberg et al., 2005).

In the delivery phase, designers monitor implementer fidelity of the curriculum, as well as provide emotional and practical support, evaluate progress, and use data to determine the intervention's ongoing viability to facilitate on-going improvement. During the post-delivery phase, fidelity data are analyzed to determine if the intervention is successful and conducive to scaling-up (Greenberg et al., 2005). Intended outcomes, such as increased student achievement, are examined in relation to FOI to determine intervention success.

When developing measurement instruments and criteria for assessing FOI, it is critical that researchers have a practical view of FOI variables. According to Lynch and O'Donnell (2005),

> Fidelity is adhering to unit and lesson purpose, goals, and objectives; adhering to unit pedagogical approaches …; following lesson sequence …; using the recommended equipment or materials …; making an adaptation to the lesson that does not change the lesson's intent. Fidelity is not reducing or modifying unit goals and objectives; reconfiguring the lesson so that your standard instructional repertoire gradually replaces parts of the new unit as originally designed; reducing the amount of behavioral change expected from participants …; varying grouping strategies outlined in the unit …; changing the unit's organizational patterns …; varying the lesson schedule …; [or] reducing the number of lessons. (pp. 9-10)

Programmatic conceptual frameworks, when available, provide structures useful for describing and organizing elements unique to an intervention to facilitate effective implementation measurement.

Conceptual Framework for the Construct of Fidelity

In 2007, Carroll et al. first proposed a conceptual framework for fidelity that is helpful in developing a plan for fidelity measurement. The framework for the fidelity construct outlined in Figure 2.1 depicts the vital FOI concepts and their relationships to one another. An examination of FOI requires that each element be measured. When examining the concept of adherence or the extent to which those accountable for implementing an intervention actually adhere to the intervention as it is designed by its developer(s), researchers typically measure two subcomponents: coverage and delivery quality. The level achieved in each subcomponent may be affected or influenced (i.e., moderated) by other variables: preexisting context, participant responsiveness, and facilitation strategies.

Adherence. Adherence refers to the extent to which an agent delivers a unit as designed or written (e.g., implementation of activities and methods) (Lynch, 2007; Lynch & O'Donnell, 2005). McGrew, Bond, Dietzen, and Salyers (1994) argued that this involves identifying program critical elements. To be considered high fidelity, Borrelli et al. (2005) recommended an 80% or greater adherence to critical elements.

Coverage. Coverage denotes delivery completeness or the content experiences that participants receive (Dusenbury et al., 2003). Also referred to as exposure or dose, this dimension may include the number of sessions, length of each session, or frequency of implementation (Dane & Schneider, 1998). Additionally, it is important to investigate whether the number of lessons and time allocated to deliver intervention content is consistent with the developer's intent (Dusenbury et al., 2003). According to Dusenbury

et al. (2003), dose can be measured by teacher logs or checklists. Century et al. (2010) asserted that *dosage* addresses the amount of time spent implementing the intervention, while *exposure* addresses the way in which the time was spent.

Delivery quality. Delivery quality is defined as the extent to which the provider approaches the developer's idealized theoretical framework in content deliverance terms (Dusenbury et al., 2003). Hamre et al. (2010) suggested that assessing delivery quality might be an important curricular intervention indicator. Dane and Schneider (1998) conceptualized delivery quality very broadly including both provider affect and expertise, which are not directly related to the prescribed content. In later studies, researchers (Century et al., 2010; Dusenbury et al., 2003; Greenberg et al., 2005; Lynch & O'Donnell, 2005; Ruiz-Primo, 2006), defined delivery quality as the proximity to the developer's theoretical implementation ideal.

Context. Context is the surrounding social structures, such as customs and cultures of groups and organizations, inter-organizational linkages, and past as well as simultaneous events, in the intervention setting (Lipsey & Cordray, 2000). Context may refer to the setting in the previous or former program, described in this research as the *preexisting context* or the context created by the new intervention, referred to as the *intervention context*.

Strategies to facilitate implementation. Strategies to facilitate implementation are support strategies used both to enhance and to standardize implementation. Such strategies include the provision of training, feedback, manuals, guidelines, and

monitoring provided to intervention implementers by developers for the purpose of enhancing FOI (Carroll et al., 2007).

Participant responsiveness. Participant responsiveness is viewed as the participants' levels of engagement in the intervention (Dusenbury et al., 2003). It refers to the degree to which the intervention stimulates interest or holds participants' attention (Durlak & DuPre, 2008). According to Carroll et al. (2007), it involves judgments made by the participants or recipients about the intervention's outcomes and relevance. The participant may be either represented as the person implementing the intervention or the person receiving the intervention. Lynch & O'Donnell (2005) defined responsiveness as the extent of student engagement consistent with the developer's intent.

An outcomes analysis can be used to identify components that are crucial to the intervention and must be implemented if the intervention is to have its anticipated effects. This assessment in turn may help update the intervention content by determining the minimum requirements for high implementation fidelity (i.e., the implementation of the essential intervention components; Carroll et al., 2007).

Measuring Fidelity of Implementation

Fidelity is the extent to which the critical components of an intended program are present when that program is enacted (Century et al., 2010). With the current emphasis on evidence-based practices, fidelity criteria development and use is an expected component of quality evaluation practices and significant with regards to treatment effectiveness (Mowbray et al., 2003). The implementation quality is thought to influence implementation effectiveness in school settings (Dusenbury et al., 2003).

Examining Fidelity Concepts

With well-developed and valid measures, fidelity can enhance statistical power in treatment-outcome studies by acting as a moderating variable in explaining outcome variance (Teague, Drake, & Ackerson, 1995; Teague, Ganju, Hornik, Johnson & McKinney, 1997) and thus reduce error in the evaluation (Greenberg et al., 2005). Practically speaking, researchers need to measure fidelity to account for negative or ambiguous findings (Hohmann & Shear, 2002) to guard against Type III errors (Patton, 1997). According to Patton (1997), a Type III error occurs when intervention evaluation does not take into account whether or not the intervention has been adequately implemented as designed, thereby attributing outcomes to an intervention that may have not been implemented or implemented as intended (Basch, Sliepcevich, Gold, Duncan, & Kolbe, 1985; Dobson & Cook, 1980; Sanchez et al., 2007). According to Greenberg et al. (2005), assessing FOI strengthens program evaluations by reducing the variance unaccounted for in the experimental design of the program or intervention. By capturing the variance explained by implementation factors, a greater unexplained variance amount is accounted for, thus, strengthening the program evaluation model.

Furthermore, FOI measures enable studies to be more standardized and conducive to replication (Mowbray et al., 2003). Failure to standardize the intervention within and between providers will inflate error variance and decrease power (Moncher & Prinz, 1991; Ruiz-Primo, 2006). Greenberg et al. (2005) referred to this replication process as diffusion because measuring implementation assists in advancing knowledge regarding best practices for replicating, maintaining, and diffusing the program or intervention.

Otherwise, program drift (i.e., deviation from the intended program) can occur during intervention or curriculum implementation and not be taken into account (Lynch, 2007; Moncher & Prinz, 1991; Mowbray et al., 2003; Resnick et al., 2005). Measuring fidelity can warn designers of possible drift. It also provides guidance in identifying and controlling unauthorized changes in an intervention (Resnick et al., 2005).

Moreover, including FOI measures in a study provides accountability to funding agencies and feedback for program improvement (Basch et al., 1985; Resnicow et al., 1998). FOI measures help explain why an intervention succeeds or fails and identify and examine changes (i.e., deletions, additions, or modifications) in the intervention that may impact outcomes (Dusenbury et al., 2003; Fullan & Pomfret, 1977; Mowbray et al., 2003). Unless FOI is examined, researchers are unable to determine whether unsuccessful outcomes reflect poorly designed interventions or failure to implement an intervention as intended (Chen, 1990; Dusenbury, Brannigan, Hansen, Walsh, & Falco, 2005).

FOI measures provide data necessary to enhance the intervention. For example, data can be used to refine the program, training manual, and observer and implementer training. Additionally, data can be used to locate issues that negate intervention effectiveness (Resnick et al., 2005; Ruiz-Primo, 2006). Examination of FOI variables also may help the researcher confirm that the control group did not implement the experimental treatment (Mills & Ragan, 2000). With a push to implement research-based interventions, Durlak (1998) pointed out that researchers should be aware that in control schools multiple interventions are often implemented without coordination. Poorly

monitored control groups further impact researchers' ability to account for student outcomes.

Likewise, examinations of FOI variables provide a mechanism for affirming the theoretical underpinnings of a program, intervention, or curriculum study (Greenberg et al., 2005). Additionally, it supports a rationale for not including data from sites that deviate extensively from the treatment model (Teague et al., 1995). Finally, FOI variables contribute to the assimilation and comprehension of the intervention's internal dynamics and actual operation (Greenberg et al., 2005). It informs the research about the kinematic phenomena (i.e., teacher-to-intervention interaction and intervention theory versus school politics) that are at play within and between the intervention and the principal agents.

It is likely, however, that all interventions are modified when implemented in practical settings. There is on-going debate between two schools of thought about FOI. One school advocates exact intervention model replication (Drake et al., 2001; Szulanski & Winter, 2002), while the other recognizes the need for adaptations to accommodate local conditions and maximize ownership (Bachrach, 1988; Domitrovich & Greenberg, 2000; Rogers, 2003). Dusenbury et al. (2003) observed that all teachers made adaptations to the lessons in their prevention study and that most of those adaptations appeared to detract from curricular objectives. Interestingly, some researchers argue that measuring fidelity is only marginally important, since adaption is necessary for an intervention to be successful in certain settings (Berman & McLaughlin, 1976). Although this school of thought recognizes the necessity of adapting interventions to

accommodate the targeted condition's unique situations and conditions, the other school of thought purports that superior outcomes occur with higher fidelity levels (Blakely et al., 1987; Elliott & Mihalic, 2004; Kam, Greenberg, & Walls, 2003). Nevertheless, failure to include fidelity measures comes with a cost: the implementation of ineffective programs or possible rejection of effective programs (Moncher & Prinz, 1991).

Methods of Measuring Fidelity of Implementation

Various direct and indirect methods and method combinations are used to measure fidelity (Dusenbury et al., 2003). Direct assessment, typically an observation, counts the frequency of occurrences of critical components to determine the percentage correctly implemented (Dusenbury et al., 2003). Critical components are the curricular or intervention elements that are considered consequential to the intervention outcomes. Indirect assessments include self-reports, interviews, and examinations of permanent products. One or more of these methods may be used for collecting data.

Direct observation. One strategy to collect data to measure FOI is on-site observations (Fullan, 1983; Mowbray et al., 2003; Resnicow et al., 1998). Previous studies have shown a link between student outcomes and fidelity using multiple observations (Resnicow et al., 1998). Fullan and Pomfret (1977) argued that direct observations are the most rigorous FOI measurement. Hansen and McNeal (1999) concurred that observations are crucial because they believe that teachers are biased in their reports.

There are, however, some disadvantages to direct observations. First, they require extensive resources and time (Fullan, 1983; Resnicow et al., 1998). Second, there are

some intervention component dimensions that are difficult to assess through observations, such as the degree of intervention philosophy, strategy, and technique comprehension (Ruiz-Primo, 2006). Third, with announced observations, validity can be limited because teachers know when an observation will occur (Resnicow et al., 1998). Ruiz-Primo (2006) recommended possible alternatives to on-site observations, such as telephone interviews, video conferencing, or videotaping.

Interview. Interviews present one of the most useful methods for assessing implementation completeness, because teachers may take more time responding in person than on a questionnaire (Resnicow et al., 1998). Nonetheless, when teachers self-report in a face-to-face interview, the interviewer should be aware that other factors, such as teacher self-consciousness, rapport, transference, and modeling could influence report validity (Paulhus & Vazire, 2007). Interviews conducted in person have the same disadvantage financially as direct observations (Fullan, 1983; Mowbray et al., 2003; O'Donnell, 2008).

Permanent products. Permanent products are examples of the participants' work during intervention implementation. Permanent product examples include student work samples, student performance assessments, and instructional session videotapes (Gresham, 1989). At times, it can be difficult to obtain permanent products limiting the assessment of tangible outcomes.

Self-report. Self-report can occur via surveys, questionnaires, logs, or checklists (Fullan, 1983; O'Donnell, 2008; Resnicow et al., 1998). Studies examining teacher self-report accuracy produced mixed results, often due to defects in study designs (Ross,

McDougall, Hogaboam-Gray, & LeSage, 2003). Self-reports can have a social desirability bias that can cause the implementer to over-report curriculum use (Resnicow et al., 1998). This is true, especially in the educational realm, if implementers suspect that the ratings may affect program funding (Mowbray et al., 2003). It also can be difficult to get participants to return surveys, questionnaires, checklists, or logs. Resnicow et al. (1998) recommended prompting participants via several contact methods, providing incentives for completion and return, and assurance that there will be no negative consequences if low implementation rates are reported.

Designing Assessments for Fidelity of Implementation

In the social sciences field, there is scant literature on FOI. Some studies reported measures taken to promote fidelity, but not to assess FOI or determine the validity or reliability of the measures (Dusenbury et al., 2003). Mowbray et al. (2003) suggested that criteria to assess fidelity consider intervention/program structure and process aspects over the entire intervention implementation.

Fidelity assessments development primarily involves one of three methods: (a) drawing from a specific model with proven efficacy, effectiveness, or acceptance; (b) gathering expert opinions (e.g., literature reviews or surveys); or (c) qualitative research to gather opinions on-site as to what works. Drawing from an effective model is the most feasible and appropriate from a conceptual and logistical perspective (Mowbray et al., 2003). Expert opinion predictive utility lessens over time in the absence of empirical findings. Additionally, experts tend to rate most components as critical (Mowbray et al., 2003).

Research conducted by the National Diffusion Network found that over half of the organizations adopting innovations modified the intervention (Rogers, 2003). Therefore, it is critical to establish effective fidelity criteria to understand better the significance and impact caused by the changes, especially those made to critical components. O'Donnell (2008) suggested identifying possible indicators or critical components of a given model, collecting data to measure the indicators, and examining the indicators' reliability and validity.

When developing critical criteria, it is important to consider the specificity levels, degree of deviation, and the amount of qualitative and empirical data required to determine fidelity (Ruiz-Primo, 2006). If elements are implemented to promote high quality FOI *a priori*, the measuring and assessing process will be more accurate and reliable. Ruiz-Primo (2006) suggested that when assessing criteria, researchers should consider adaptive planning, tailored training, sufficient sample size, local material development, teacher training, program characteristics, teacher characteristics, and organizational characteristics. Mowbray et al. (2003) suggested assessing the fidelity criteria reliability and validity by: (a) calculating the inter-rater agreement range on ratings and scores, (b) examining differences in fidelity scores (i.e., divergent validity) across program types expected to be different, (c) examining the convergent validity or agreement between at least two different data sources measuring the same thing, (d) examining the internal data structure empirically and in relation to expected results, and (e) measuring the relationship between fidelity measures and expected outcomes for participants.

In the education field, there is empirical evidence upon which to base assessment measures for fidelity. Assessing fidelity is an on-going process because opinions about critical components are always changing, as are political, economic, and clinical circumstances impacting the study (Mowbray et al., 2003). Adherence and dosage are two fidelity dimensions that dominate current fidelity literature and provide critical information about the faithful implementation of interventions. In the next section, these and other major prominent factors affecting FOI will be discussed.

Institutional Barriers and Opportunities – Factors Affecting Fidelity of Implementation

Before focusing on factors that affect FOI, it is important to understand how the fidelity framework fits into a larger improvement model that focuses on using instructional materials as a key intervention. In this model, designed by Century et al. (2010), an instructional intervention is designed and implemented with the expectation that, if the teacher implements the intervention as originally envisioned by the designer, the desired student outcomes will occur. Implementation, however, does not always manifest itself as envisioned. According to *What Works Clearinghouse* (U.S. Department of Education, 2011), there are several factors in intervention research design that can improve findings and, as such, should be heeded when developing implementation criteria including: using random assignment, collecting and reporting FOI assessment, and using standardized measures. Unfortunately, to date, no standardized measures have been developed to measure the factors influencing FOI (Lynch, 2007). Ruiz-Primo (2006) divided the influential FOI factors into two groups: intervention context

characteristics (e.g., theoretical agreement, satisfaction levels, and intervention effectiveness perceptions) and intervention characteristics (e.g., complexity, time, resources, and manual quality).

The conglomeration of factors that influences FOI is two-pronged: non-school-based and school-based. Although non-school-based factors can be contextual or programmatic, school-based factors are typically contextual. Non-school-based factors can include any factor outside the school or organizational boundaries where the intervention is implemented, such as intervention design, compatibility with intervention sites, politics (e.g., board of education), budget, professional organizations, and overall fit. Internal factors can include principal actors' characteristics (e.g., gender, title, and experience level), perceptions (e.g., end of grade testing and curricular choices), and expectations (e.g., student success and administrator support), in addition to school culture and practices that occur within the school or organization boundaries.

In the educational realm, factors influencing implementation are intertwined, forming a nested influence setting. In other words, layers of agents and context are nested within the setting. For example, the teacher, students, classroom environment, others in the school, school culture, parents and community, and to an extent, the school district, state and federal authorities, and policies all may influence classroom events. Generally speaking, the preexisting social, historical, and economic contexts in which an intervention is implemented (Datnow & Castellano, 2000) influence FOI because schools are not entities removed from the environment (Kurki et al., 2006). Because these ever-

changing and adapting multi-environmental entities are influenced by human beings they are subject to an array of factors that influence the FOI level.

Non-School-Based Factors

Non-school-based factors influencing FOI can be contextual or programmatic. Preexisting contextual factors include issues relating to policies/politics, district and state budgets, district fit, partnerships, and introduction of new innovations or curricula. At the local level, policies and politics relating to budget changes, introduction of new tests or curriculum guidelines, elected local officials (Datnow & Castellano, 2000) and financial support for effective training (Everhart & Wandersman, 2000) may influence FOI. Brunette et al. (2008) stated that mandates and funding priorities can further impact FOI. At the district level, innovations already in place and the preferences or priorities established by central administrative office can predispose a certain FOI level (Durlak & DuPre, 2008). Many policies are not initiated at the local level, but at the federal and state levels, thus ramifications trickle down impacting FOI. Partnerships, the lack of partnerships, and partnership type contribute to the FOI level in that partnerships can provide resources and community buy-in that often are necessary to implement new interventions (Century et al., 2010; Datnow & Castellano, 2000; Kurki, Boyle, & Aladjem, 2006; Riley, Taylor, & Elliott, 2003). Theoretical fit between the intervention and adopters' philosophies and contexts increase the likelihood that adopters will implement the curriculum as envisioned by the developer (Moncher & Prinz, 1991).

School-Based Factors

School-based factors typically influence FOI within the school's boundaries. School-based contextual factors are influenced by implementer perceptions. These factors may be physical, mental, or ideological. Additionally, numerous variables or variable combinations interact within complex school settings. School-based factors can be sub-divided into classroom factors and non-classroom factors.

Classroom factors. Inside the classroom, there are three possible influences on FOI: (a) the teacher, (b) the classroom climate, and (c) peer relations. Teachers are considered by most policymakers and school change experts to be the heart of the educational change process (Datnow & Castellano, 2000). Teacher characteristics that contribute to FOI levels are identical to characteristics of teacher effectiveness. For example, FOI may be influenced by the teacher's age, career stage, race, culture (Datnow & Castellano, 2000), educational attainment (Goldhaber & Brewer, 1996; Huang & Moon, 2008), teacher certification (Darling-Hammond, Berry, & Thomerson, 2001; Goldhaber & Brewer, 1996; Palardy & Rumberger, 2008), skill proficiency (Durlak & DuPre, 2008; Tyack & Cuban, 1995), and experience level (Huang & Moon, 2008; Rivkin, Hanushek, & Kain, 2005). Rohrbach, Graham, and Hansen (1993) found high FOI associated with providers who had fewer years of teaching experience, high implementation self-efficacy, positive program acceptance, and an intervention compatible teaching style. In contrast, Young et al. (1990) found the opposite when examining the role of teacher experience in FOI. In summary, it is clear that teacher

characteristics are critically important in fidelity and should be monitored and assessed as a key measure of classroom receptiveness.

Other implementer characteristics that influence FOI involve implementer philosophy, perceptions, and expectations. Using 23 contextual factors identified in a meta-analysis to influence fidelity, Durlak and DuPre (2008) pointed out three that were related to implementer philosophy, perceptions, and expectations. For example, the extent to which a teacher feels a need for or sees the benefit of an intervention influences how thoroughly the intervention is implemented. Everhart and Wandersman (2000) examined implementers' perceived success and expected continuation of the program as determinants of FOI. They concluded that teacher commitment and self-efficacy influence implementation.

Additionally, the need for ownership in the change process (Datnow & Castellano, 2000) and intervention prioritization (Brunette et al., 2008) shape teachers' actions, which in turn influences FOI. Berman and McLaughlin (1976) pointed out that principals' use of active support increases teacher change possibilities and perceived intervention success, thus enhancing chances of higher FOI levels. Additionally, concurrence between the intervention and implementer philosophy on key program elements greatly influences program acceptance and implementation fidelity. Likewise, when change rhetoric associated with the intervention matches the realities of teacher experiences (Datnow & Castellano, 2000), higher FOI levels occur.

Finally, according to Datnow and Castellano (2000), teachers' engagement levels in the intervention were influenced by their beliefs, practices, and the quality of the

relationships among teachers and students within the school culture. Discipline-specific teachers must have similar teaching philosophies for a discipline-wide intervention implementation to be successful. Other teachers at the school not implementing the intervention must be cognizant and supportive to minimize implementation disruptions and maximize teacher-to-teacher and teacher-to-student morale.

Greenberg et al. (2005) argued that teachers play a primary role in promoting positive climate and positive peer relations within the classroom, which in turn influences positive FOI levels. Classroom climate is a social and psychological aggregation that can influence FOI positively or negatively, reflecting the teacher's classroom management techniques, decisions to empower students to participate in decision making, and appropriate social and emotional modeling.

Non-classroom factors. Factors related to school but outside the classroom, often referred to as the school's organizational capacity influence external organizational relationships. The influence, however, does not have the same immediate or direct impact as do teachers. Organizational predispositions and practices can explain, in part, FOI levels (Riley et al., 2003).

An organization's capacity is defined as its potential to perform or its ability to apply skills and resources to accomplish stated goals. At the center of organizational capacity is governance and leadership. Like the characteristics of effective teachers, specific characteristics of effective administrators are indicative of higher FOI levels. For example, according to Leithwood, Day, Sammons, Harris, and Hopkins (2006), an effective administrator sets the direction, develops professional relationships between

teachers, redesigns the organization to maximize instructional time, and manages instruction to develop an effective school. As such, an administrator is connected with all the organizational elements that influence FOI, either directly or indirectly. In physical education, for example, limited instructional time has grown to be a defining, negative physical education program characteristic, limiting intervention effectiveness. To a certain degree, school administrators control school scheduling. Lack of instructional time serves as a barrier to effective implementation (Century et al., 2010) and can be connected with a leader's ineffective time and space usage, operating procedures, and task assignments. Time constraints often lead teachers to feel pressured and unable to implement the curriculum as designed (Borman et al., 2007; Klinger, Ahwee, Pilonieta, & Mendez, 2003; Klinger, Cramer, & Harry, 2006). Interventions cannot be implemented with fidelity if allotted instructional time does not equal or exceed the intervention exposure requirements. Therefore, a leader who is an effective time manager is more likely to positively impact implementation than a less effective one. When the internal and external environments are healthy, the leader is able to focus on securing resources and modifying the structure more effectively, thus positively influencing levels of FOI in general.

Programmatic or Intervention Factors

Programmatic factors are directly related to the curriculum or intervention, itself. They may include the design, complexity, and intervention quality, in addition to time required for implementation, materials supply, and the observers' and implementers' training quality (Gresham, 1989; Ruiz-Primo, 2006). Complexity level can refer to the

number of interrelated components, number of steps involved, program comprehension, and its implementation precision (Ruiz-Primo, 2006). Typically, the higher the complexity level, the lower the FOI level. Intervention characteristics (Berman & McLaughlin, 1976; Durlak & DuPre, 2008, Kurki et al., 2006) can influence FOI through the educational treatment or technology, resource levels, proposed scope of change, and implementation strategies and requirements. The intervention adaptability (flexibility) and compatibility (contextual appropriateness or fit) also influence FOI (Durlak & DuPre, 2008; Kurki et al., 2006). The program design, if not theoretically based and tested for design specification fidelity, can lead to intervention failure (Greenberg et al., 2005).

After reviewing the relevant FOI barriers/opportunities literature, teachers, instructional time, and intervention complexity appear to have the greatest potential impact on FOI. In the next section, school-based physical activity interventions will be discussed. Their complexities (e.g., number of components and research design) will be reviewed and outcomes summarized.

School-Based Physical Activity Interventions

Multi-activity physical education curricular models continue to dominate school-based physical education programs (Siedentop, Mand, & Taggart, 1986). Although a number of scholars have questioned the effectiveness of multi-activity physical education (Ennis, 1999; Locke, 1992), curriculum developers rarely have successfully challenged the anti-social gymnasium culture created, in part, by the multi-activity physical education model (Ennis, 1995). Initial stabilizing conditions, such as the emphasis on sports and games popular to a region, short 2- or 3-week units, and teachers' beliefs that

skills should be taught first and games played second, are historical characteristics of multi-activity models (Lund, 1993) that act as strong attractors to sustain the multi-activity curriculum. These initial conditions lead to attractor states (i.e., lack of instruction, lack of teacher and student accountability, and marginalization of low-skilled students) resistant to change in ineffective K-12 physical education programs (Parker & Curtner-Smith, 2005). Scholars and researchers have responded to the multi-activity curriculum challenge (i.e., ineffective attractor states) by creating bifurcations in the form of large-scale physical activity interventions. They have attempted to implement them in multi-activity physical education programs with varying results. Examples of these attempts are summarized in the next few paragraphs.

Because elementary physical education often is taught by a single teacher in a small, relatively isolated environment, physical activity interventions frequently are initiated at this school level. Elementary interventions have been designed to increase physical activity levels, improve fitness levels, and prevent and/or treat childhood overweight and obesity. Specifically, interventions emphasize increasing physical activity during physical education, cognition and classroom-based knowledge, and physical activity during non-school hours. Principal outcomes include time engaged in physical activity, physical activity and health behaviors knowledge, cardiovascular risk factors (e.g., cholesterol), and body composition (e.g., body mass index and body fat).

Physical education-based physical activity interventions are characterized by diversity in structure and program. For example, multiple, varying length (8 weeks to 2 years), physical education-based physical activity interventions have resulted in higher

amounts of moderate to vigorous physical activity during physical education (Donnelly et al., 1996; Luepker et al., 1996; Sallis et al., 1997; Simons-Morton, Parcel, Baranowski, Forthofer, & O'Hara, 1991) and increases in daily physical activity (Harrell et al., 1998; Harrell et al., 1996). Most interventions, however, have not produced significant changes in body composition (Caballero et al., 2003; Donnelly et al., 1996; Luepker et al., 1996). Accordingly, interventions have rarely resulted in improvements in fitness, adiposity, and/or cardiovascular risk factors. Although, significant increases in physical activity have been detected, they may not be adequate to promote body composition and/or fitness improvements. Additionally, the physical activity levels may not have been of adequate intensity, duration, or type (i.e., aerobic activities) to facilitate fitness improvements.

Elementary School Interventions

Several elementary level interventions have demonstrated promise in changing elementary physical education to promote moderate to vigorous physical activity and increased fitness understanding. The *Sports, Play, and Active Recreation for Kids (SPARK)* project, for example, was a 2-year quasi-experimental intervention utilizing fourth- and fifth-grade students from seven California elementary schools (Sallis et al., 1997). The randomized controlled trial incorporated three teacher conditions: classes taught by physical education specialists, trained classroom teachers, and untrained classroom teachers. Findings indicated minutes per week of moderate to vigorous physical activity during physical education were 40 min in the physical education specialist-led schools, 33 min in the classroom teacher-led schools, and 18 min in the

control schools. There were no significant differences, however, between the three treatments for out-of-school physical activity or changes in body composition. Furthermore, females increased their scores on fitness measures, while males did not. These two later points are important because they may suggest that school-based physical activity interventions based on a "one-size-fits-all" approach may not adequately address the intricacies of the complex adaptive systems represented by schools and physical education.

An elementary level intervention entitled, The *Child and Adolescent Trial for Cardiovascular Health* (renamed *Coordinated Approach to Child Health [CATCH]* in 1999) provided additional insights into programmatic factors impacting school-based moderate to vigorous physical activity. This curriculum was evaluated in a 3-year randomized controlled trial implemented in third grade at 96 elementary schools in California, Louisiana, Minnesota, and Texas in an attempt to improve students' diet and physical activity levels. *CATCH* utilized school-based (i.e., physical education, classroom curricula, and food services) and family-based (i.e., family fun nights and home curricula) modules in the experimental schools and traditional physical education and food services in the control schools (Luepker et al., 1996). Experimental school cafeterias served foods lower in fat and sodium, physical education teachers increased moderate to vigorous physical activity to at least 50% of class time, and schools established non-tobacco use policies. As a result of the *CATCH* curriculum, the time expended in moderate to vigorous physical activity during physical education improved significantly in the experimental schools compared to the control schools. Students in the

experimental schools also engaged in more vigorous physical activity compared to students in the control schools. During the 3-year follow-up, participants from the experimental schools experienced significantly larger physical activity amounts compared to the participants from the control schools, although the differences were smaller than those reported during the intervention (Nader et al., 1999).

A 5-year, federally funded school-based intervention called *Science, PE, & Me!* (Sun, Chen, Zhu, & Ennis, 2012) was evaluated from 2003 to 2008 in a large, urban school district in Maryland. The intervention utilized a randomized clinical trial to design, implement, and evaluate a constructivist-based, science-enriched, health-related physical education curriculum (30 unique lessons per grade) for Grades 3-5. Both the experimental group and control group contained over 6,000 third- through fifth-grade students from 30 elementary schools. Results from the intervention indicated that students in the experimental group were moderately physically active in most lessons (Chen, Martin, Sun, & Ennis, 2007) and increased their fitness-based knowledge (Sun et al., 2012).

Two elementary school-based physical activity interventions implemented outside the United States also are worth noting. First, a 6-month overweight prevention study incorporating physical activity and nutrition components targeted elementary schools in different areas of Chile (Kain et al., 2004). The physical activity intervention portion included classroom-based lessons in Grades 1-8, 90 physical activity min per week in Grades 3-8, and active recess. The results revealed significant improvements in fitness, body mass index, and waist circumference in the experimental schools compared to the

control schools. This intervention offers support that the school culture can be adapted to provide larger physical activity quantities and better-quality nutrition to promote changes in fitness and body composition.

Secondly, Manios, Moschandreas, Hatzis, and Kafatos (1999) conducted a 6-year intervention study in Greece targeting students in Grades 1-6. The program included educational seminars for parents, and enhanced physical education (i.e., two 45-min physical education classes per week) and classroom education for the children. After 3 years, children in the experimental schools significantly improved their health knowledge and reported more leisure-time physical activity compared with students in control schools who received traditional physical education. Long-term school-based interventions are challenging to implement due to a multitude of factors. Longer interventions are more complex, which in turn negatively impacts implementation fidelity. This study, along with *Science, PE, Me!*, provided evidence that long-term school-based interventions can be implemented with success and value to stakeholders.

Some elementary school-based studies have targeted the after-school hours as the primary intervention focus. For example, a study (Owens et al., 1999) conducted in Georgia focused on overweight 9- to 11-year-old children. Children ($N = 81$) were recruited from local elementary schools and transported after school to the Medical College of Georgia. In this cross-over design, participants were randomly assigned to one of two conditions: an exercise/no exercise or a reversed no exercise/exercise format. Each group participated for 8 months in their condition: equally divided into 4-month sub-conditions. Physical activity was offered 5 days per week for 40 min per session and

elicited heart rates at or above 150 beats per minute in the participants. Children who exercised during the first 4 months had a reduction in body fat during the first 4 months compared to the participants who did not exercise during the first 4 months. When the initial exercise group stopped exercising after 4 months, their body fat levels returned to pre-intervention levels (Owens et al., 1999).

In a separate study conducted by The Medical College of Georgia Exercise Project group (Barbeau et al., 2007), researchers recruited approximately 300 African American females in Grades 3-5 for a 10-month, after-school physical activity randomized controlled trial. The 80-min physical activity program (25 min for skills, 35 min for moderate to vigorous physical activity, and 20 min for toning), based at the participants' school gymnasiums, included ego-oriented and mastery-oriented physical activity. Initial results indicated that participants in the physical activity experimental group significantly (all p values $< .05$) increased their moderate to vigorous physical activity levels and demonstrated improvements in cardiovascular fitness (measured by a graded treadmill test) and body composition (measured by DXA) compared to participants in the control group (Barbeau et al., 2007).

The latter two physical activity interventions suggested that school-based interventions should be convenient to and for the participants. One intervention provided transportation and the other intervention utilized local school property, which in most cases was convenient for the participants. Participant responsiveness and engagement is one of the six components to consider when measuring intervention FOI.

Middle School Interventions

Several middle-school physical activity interventions produced positive effects on fitness, physical activity levels, and adiposity (Alexandrov, Maslennikova, Kulikov, Propirnij, & Perova, 1992; Colchico, Zybert, & Basch, 2000; Hoerr, Nelson, & Essex-Sorlie, 1988; Vandongen et al., 1995).

The *Cardiovascular Health in Children and Youth* Study (*CHIC II*; McMurray et al., 2002), for instance, was a randomized controlled study conducted over an 8 week period. Four middle schools enrolling 1,140 students participated in the study with each school assigned to one of the following conditions: physical activity only, education only, education and physical activity combined, or control. The physical activity middle schools received aerobically-based physical education for 30 min, 3 days per week. The education middle schools conveyed health content in a classroom setting by a certified health education teacher, while the control group followed their usual health and physical education curriculum. The intervention produced a significant increase in cardiovascular fitness measured by a cycle ergometer test in the education and physical activity combined group compared to the education only group. No significant cardiovascular fitness changes occurred between the other groups. In addition, at study's conclusion, systolic blood pressure was significantly lower in all three interventions groups compared to the control group.

The Class of 1989 Study, conducted as part of the *Minnesota Heart Health Program*, was a 7-year cross-sectional and longitudinal randomized controlled intervention enrolling approximately 2,300 students in Grades 6-12 (Kelder, Perry, &

Klepp, 1993). During sixth and seventh grades, students completed an annual physical activity behavior survey. The initial physical activity intervention implemented in eighth grade encouraged participants to engage in physical activity outside of school. The next physical activity intervention, implemented in 10th grade, was a peer-led curriculum focused on promoting regular physical activity and healthy eating. During follow-up, intervention group females self-reported higher physical activity levels compared to the control group females. Males showed no significant difference, supporting the premise that physical activity interventions may be more effective if they contain gender-specific components.

The *Middle School Physical Activity and Nutrition* (*M-SPAN*; Sallis et al., 2003) intervention was a randomized controlled 2-year project that focused on revising school policy and environment to impact both physical activity and nutrition components. The physical activity component focused on increasing physical activity levels during physical education and making school climate revisions to facilitate increased physical activity during leisure time (i.e., before school, at lunch, and after school). Researchers worked cooperatively with a school-based health policy committee composed of school administrators, staff, parents, and students to develop and enact school policy changes. Male students in the intervention schools demonstrated a significant increase in physical activity during physical education and school-leisure time compared to the males in the control schools. The findings from this intervention indicated that changes to school-based policies can improve intervention FOI and positively impact intervention outcomes.

The purpose of the 3-year *Trial for Adolescent Activity Girls* (*TAAG*; Webber et al., 2008) was to change physical education and health education through teacher training, school health services, school environment, faculty/staff health promotion, and school-community linkages. Researchers employed a randomized controlled trial to measure sixth- through eighth-grade females' changes in physical activity levels using self-report inventories and accelerometers. The findings showed no significant changes in physical activity levels after the 2-year, staff-directed intervention. The 3rd-year component, *TAAG Program Champion* (Webber et al., 2008), however, indicated females in the experimental schools were significantly more physically active than females in the control schools. *TAAG Program Champion* was facilitated by school-based teachers and staff recruited and trained by *TAAG* intervention staff to sustain the intervention. The 3rd-year findings supported the direct relationship between implementer ownership, delivery quality, and FOI level.

Planet Health was a 2-year experimental intervention using 10 randomly matched middle schools and approximately 1,500 students (Gortmaker et al., 1999). The intervention entailed both a physical activity and nutrition component taught in both physical education and classroom settings, while control schools continued teaching their traditional curriculum. The physical activity component was designed to increase moderate to vigorous physical activity and reduce television viewing time. The females in the experimental condition demonstrated a significant improvement in body composition compared to the females in the control group. Both males and females in the

intervention group reported significantly less television viewing compared to the control group. No significant differences were reported for physical activity levels.

Active Winners was a multifaceted intervention conducted in South Carolina that utilized school, after-school, home, summer camp, and community components (Pate et al., 2003). The intervention began with a summer camp (5 hr per day) focused on fitness, sport skills, academics, and social skills for children who just completed Grade 5. The summer camp was followed by physical activity interventions in Grade 6, implemented both during and after-school, another summer camp after completing Grade 6, and a one semester follow-up in Grade 7. The home and community portions of the intervention were implemented throughout the study duration, and included educational newsletters, take-home physical activities, and family nights. Some components (i.e., home, school, and community) were not implemented as designed due to minimal resources. This may have contributed to the finding of no significant differences in self-reported physical activity between the experimental and control groups.

High School Interventions

Comparatively speaking, although there are fewer published school-based physical activity intervention studies targeting high school students, a majority of these do include a follow-up period; an important component missing in many school-based physical activity interventions.

For example, *Project FAB* was a 4-month intervention trial focused on females (n = 58) in Grades 10 and 11 from two secondary schools (Jamner, Spruijt-Metz, Bassin, & Cooper, 2004). The enriched physical education program for females incorporated

various physical activities for 60 min per day for 4 days per week and one class per week dedicated to strategies for becoming more active and physical activity health benefits. The females in the experimental school showed significant improvements in cardiovascular fitness (as measured by a cycle ergometer test) and self-reported moderate to vigorous physical activity compared to the control school at intervention completion.

New Moves was a 16-week, randomized controlled intervention with an 8-week follow-up delivered in Grades 9 and 10 for females from six secondary schools in the Minnesota Twin Cities area (Neumark-Sztainer, Story, Hannan, & Rex, 2003). The major intervention component was enriched physical education offered 4 days per week. Over the four physical education days per week, 1 day was reserved for a guest instructor from the community, 1 day was reserved for resistance training, and 2 days was reserved for student-selected activities. The control schools continued to teach their usual physical education curriculum. The participants' body mass indices and self-reported physical activity levels displayed no significant differences between experimental and control groups. Students' physical activity readiness also was recorded in the intervention. Although no significant differences in readiness scores were detected between the two groups, during follow-up the intervention group females' readiness scores were significantly different from scores of the control group during follow-up. The later result is important, as it verifies that follow-up periods can provide valuable information about intervention success.

The *Lifestyle Education for Activity Program* (*LEAP*) was a 2-year randomized controlled project conducted with ninth-grade females from 24 secondary schools in

South Carolina (Pate et al., 2005). The intervention contained two modules intended to create a climate that encouraged physical activity in high-school females. Module 1 included a 1-year physical education program focused on developing motor skills, encouragement, modeling success, and moderate physical activity. Module 2 included behavior and health education promoting lifelong physical activity. After the first year, vigorous physical activity time was significantly higher in the intervention schools than in the control schools, with even higher vigorous physical activity time in high fidelity schools. Additionally, Saunders et al. (2006) reported a significant linear dose-response, indicating a dose effect in vigorous physical activity from low to high FOI schools. *LEAP* is one of the first school-based physical activity interventions to report implementation fidelity. FOI levels were determined using a 25-item lesson observation checklist, a *LEAP* criteria scale, and school administrator interviews. Schools were ranked ordered using these data sources. Schools consistently ranked in the bottom third of the measures were labeled as low fidelity schools.

Summary

To date, even though some school-based physical activity interventions have provided mixed results, both positive and negative findings can inform researchers when designing new interventions. Interventions that included implementer training, support from intervention staff, and philosophically matched implementer ideals consistently produced significant results. Interventions involving follow-up periods also tended to produce significant outcomes. Although few interventions measured FOI, the ones that did suggested that FOI can help explain variance between outcomes.

Limitations of School-Based Physical Activity Interventions

There are several limitations accompanying the school-based physical activity interventions published to date. In large part, the intervention trials used self-report physical activity measures and fitness measures administered by individuals with a wide range of experiences and qualifications. Although self-report physical activity measures are common, reliability and validity issues surround the accuracy and honesty of recall-based measures, especially when completed by children and adolescents. Further, although the fitness measures used in the intervention studies were validated as age-appropriate, test administrators varied in their ability to consistently implement the tests as validated.

Clearly, long time periods and specific protocols are required to invoke educational, behavioral, and physiological changes. Unfortunately, many studies reviewed were short in duration and/or failed to include vigorous physical activity. Improvements in fitness levels generally occur only as a result of high-intensity, long-term, long-duration, specific physical activity training. It may be unrealistic to anticipate significant improvements in fitness without an appropriate training environment.

Many intervention studies used physical education as the vehicle to deliver physical activity. On the surface, it is logical for school-based physical activity interventions to target physical education. Because students (and researchers) assume students are going to be active in physical education, researchers reason that inserting a physical activity intervention into an already physically active class will increase students' physical activity and fitness levels, while causing the least interruption in daily

routine. However, given the limited instructional time allocated to physical education, students spend a relatively small portion of their school week in physical education. Physical activity time in physical education is constrained and can only be as high as the time the school allocates for physical education.

Given the interventions examined in this review, it is clear that physical activity levels in schools can be increased. Further, follow-up studies and programs to sustain the increased physical activity levels after intervention completion is gaining value because teachers often revert to the previous program. Long-term focus on the intervention appears necessary to combat the tendency of children and adolescents to gain nonessential body fat (Ogden, Carroll, Kit, & Flegal, 2012). Using the above interventions as prototypes, intervention success during implementation and during follow-up appears to depend primarily on the fidelity levels achieved by intervention implementers (i.e., physical education teachers, classroom teachers, and after-school physical activity directors), physical activity levels embedded in the lessons, inclusion of relevant knowledge components, and the level of intervention complexity. Intervention complexity impacts implementers' abilities to be faithful to the intervention. Adherence, coverage, and delivery quality appear to be key indicators for FOI. These three components impact participant responsiveness and intervention context, and as a result, impacts outcome variables, such as student physical activity levels and knowledge change. These critical elements define intervention success and provide an opportunity to better understand the complex adaptive systems known as schools. Although schools have long resisted change, intervention efforts to bifurcate these stable systems and

generate reform rely on clear understandings of the systems' initial conditions and attractor states, and the intervention itself needs to be designed and implemented with fidelity.

CHAPTER III

RESEARCH METHODS

The modified conceptual framework for implementation fidelity presented in chapter II (Figure 2.1) includes the components of implementation fidelity (RQ1), intervention outcomes (RQ2), and moderating factors that may impact the degree of fidelity (RQ2 and RQ3). The conceptual framework suggests that fidelity is influenced by the moderating factors of participant responsiveness to a program, facilitation strategies, and context. Furthermore, although the relationship between an intervention and its outcomes is external to implementation fidelity, the degree of implementation fidelity achieved can affect this relationship. This study examined implementation fidelity in four middle schools implementing the *Science of Healthful Living* (*SHL*) curriculum. This research design to examine fidelity of implementation (FOI) was conducted within the overall structure of the NIH sponsored randomized clinical trial currently occurring to investigate outcomes associated with the *SHL* curriculum. Research questions that guided this study were:

1. To what extent do teachers implement lessons from a sixth-grade physical education curricular intervention with fidelity (adherence – coverage and delivery quality)?
2. What is the relationship between implementation fidelity and student outcomes (i.e., knowledge gain and physical activity intensity levels)?

3. What reasons do teachers offer for making changes to the intended curriculum as they implement it in their school contexts (moderating variable of context)?

In this chapter I will first discuss the relevance of the research questions within a FOI framework. I will then discuss the *SHL* project in general, emphasizing the unique characteristics of my research design, research sites and participants, research sampling, variables and measures, research procedures (i.e., ethics, data collection, data reduction, and data analysis), threats to validity and reliability, credibility, and my role as the researcher.

Fidelity of Implementation Framework

Fidelity Variables – Adherence (RQ1)

In the modified conceptual framework used in this research, the measurement of implementation fidelity is limited primarily to the measurement of adherence (RQ1), with its subcomponents – coverage and delivery quality. Thus, adherence relates to the coverage of the intervention; specifically whether the key components of the intervention have been received by the participants as frequently and for as long as planned in a high quality manner. In this study, adherence was used as the principal measure of fidelity and included the subcomponents of coverage and delivery quality.

Outcome Measures (RQ2)

Student physical activity intensity levels and student knowledge growth were the intervention outcomes in this study. The level of implementation fidelity is expected to influence intervention outcomes. The dilution of intervention components and/or deviation from the prescribed curriculum can have unintended consequences on physical

activity intensity levels and student knowledge growth outcomes (Dane & Schneider, 1998) in the *SHL* intervention.

Moderating Variables for Fidelity of Implementation (RQ2 and RQ3 – context only)

Teacher satisfaction. One critical moderating variable in the examination of implementation fidelity is participant responsiveness, or how well participants respond to, or are engaged by an intervention. It involves decisions by participants about the outcomes and applicability of an intervention (Carroll et al., 2007). Responsiveness refers both to individuals responsible for delivering it and individuals receiving the intervention. In this study, participant responsiveness was the teachers' satisfaction level with the intervention components.

Intervention facilitation strategies. Complex interventions that are clearly described are assumed to be easier to implement with high fidelity than vague or poorly defined interventions (Greenhalgh, Robert, Macfarlane, Bate, & Kyriakidou, 2004). Thus, a second moderating variable in this research was the use of adequate facilitation strategies to increase opportunities for higher and more standardized fidelity. In this study, facilitation strategies were defined as the number of professional development sessions each participating teacher attended during the study and the number of coaching visits to each sixth-grade physical education class.

Preexisting context. Lipsey and Cordray (2000) emphasized the importance of taking into account a third moderating factor, preexisting context, in examining the quality of implementation fidelity for the new intervention. Preexisting context includes preexisting social systems, such as inter-organizational linkages, structures and cultures

of organizations and groups, as well as concurrent events. For example, the professional relationship between a school's administration and its physical education faculty can impact physical education instructional time. The amount of instructional time afforded to physical education teachers plays a deciding role in what can and cannot be taught and at what level of detail. In this study, preexisting context was examined through the teachers' positive and negative assessments about the appropriateness of implementing the intervention at their school and the context-related reasons they cite for making modifications to the intervention.

Carroll et al. (2007) suggested that there are complex relationships at work between the moderators that may further influence the relationship between an intervention and its implementation fidelity. For example, fewer minutes of teacher training may be supplemented with increased opportunities for teachers to receive coaching. Likewise, very well designed teacher performance evaluations may compensate for fewer training and coaching experiences. In summary, the conceptual framework for the implementation fidelity construct suggests that different moderating factors might enhance or constrain the implementation process and its fidelity. These factors often interact, and the impact of one factor on implementation fidelity might be affected by another moderating factor. The conceptual framework suggests that all factors should be assessed systematically when performing a process evaluation study. In this study, the conceptual framework in Figure 2.1 was used as a theoretical model to organize data collection and analyses.

The purpose of this study was to examine physical education teachers' FOI of a middle school (Grades 6-8) *SHL* physical education curricular intervention. In this investigation, the reasons teachers cited for making adaptations to the intervention and the impact FOI had on student physical activity intensity levels and knowledge growth were examined as elements of implementation fidelity. To this end, I collected data via instruments and procedures that support: (a) qualitative analysis of adherence and its subcategories of coverage and delivery quality (RQ1), (b) quantitative analysis of the impact that FOI had on student physical activity intensity levels and knowledge growth (RQ2), and (c) qualitative analysis of teachers' rationale for making modifications to the intervention (RQ3).

Science of Healthful Living **Project**

I implemented the dissertation study during the second year of the larger, 5-year *SHL* project conducted by the University of North Carolina at Greensboro with support provided by the National Institutes of Health. The goal of the *SHL* project was to design and evaluate a science-based middle school *SHL* curriculum to increase students' cognitive knowledge and interest in health-related science, increase their intention to pursue a life science-related career, and improve their family members' understanding of National Institutes of Health funded clinical and basic research results. The *SHL* project designed a health-related science-based curriculum for middle school physical education. Consisting of two, 20-lesson units, *Cardio Fitness Club* and *Healthy Lifestyles*, the lessons engage students in the scientific inquiry process in physical education. Students engage in physically active lessons to examine the effects of exercise and healthy

nutrition on their bodies. They explore a range of topics such as the physiology of exercise, food-fueled energy systems, caloric balance, nutrition/portion sizes, and the role of physical activity and nutrition in stress management. The scientific inquiry process is embedded in each lesson using the 5 Es (i.e., Engagement, Exploration, Explanation, Elaboration, and Evaluation) learning cycle strategy (Bybee et al., 1989) to assist physical educators and students engage in the scientific inquiry process in a hands-on approach to science. Students are afforded opportunities to make predictions about physical activity concepts, test their predictions, and draw conclusions about their predictions in physically active environments. Additional information about the larger *SHL* project can be found at http://www.uncg.edu/hhs/science-of-healthful-living.

During Year 2, the *SHL* project was set in seven local education agencies in the Piedmont region of North Carolina. Twenty-five middle schools, including five new middle schools recruited in Year 2 of the study, participated in the project, including over 16,000 diverse students and 70 middle school physical education teachers. Using a randomized controlled experimental design, the middle schools were stratified during grant Year 1 into matched pairs according to socioeconomic status (derived from free and reduced-price meals school-level data) and scores on end of grade tests in mathematics and reading. One school in each pair was randomly assigned to teach the *SHL* curriculum (experimental school) and one school taught the traditional sport-based or multi-activity physical education (control school).

Research Design

I used a mixed methods design to collect and analyze both quantitative and qualitative data examining implementation fidelity in the *SHL* project (Creswell, 2013; Creswell & Plano Clark, 2011; Tashakkori & Teddlie, 2010). Using multiple data collection strategies provided a more complete analysis than any single method alone could accomplish when addressing the problem (Creswell, 2013; Creswell & Plano Clark, 2011; Tashakkori & Teddlie, 1998; 2010). A mixed methods design was appropriate because FOI measurement and analysis required both quantitative and qualitative data collection and analysis during various study stages. Identifying emergent and unanticipated practice behaviors in the *SHL* project, such as the reason(s) why a teacher decides to omit a critical lesson component (i.e., the 5 Es), may have been difficult if predetermined quantitative measures were used in isolation. Mixed methods approaches, however, offered complementary insights into understanding the faithful implementation of the *SHL* curriculum (RQ1), the impact FOI had on student outcomes (RQ2), and the reasons teachers cited for making modifications to the intended curriculum (RQ3).

Quantitative Method

Multiple regression was used to measure the relationship of the predictor variables to the criterion variable. According to Field (2009), "Regression analysis is a way of predicting an outcome variable from one predictor variable (simple regression) or several predictor variables (multiple regression)" (p. 198). Specifically, the simultaneous method (i.e., forced entry) of multiple regression allows the researcher to specify the set

of predictor variables that make up a multiple regression model and control for a relatively low number of cases (Field, 2009). The success of this model in predicting the criterion variable is then assessed.

Qualitative Method

Qualitative research is "an inquiry process of understanding …" in which "the researcher builds a complex, holistic picture; analyzes words; reports detailed views of informants; and conducts the study in a natural setting" (Creswell, 2013, p. 300). In this approach, the researcher makes knowledge claims based on the advocacy/participatory (Mertens, 2003) or constructivist (Guba & Lincoln, 1982) perspectives. In qualitative research, data are collected from those exposed to the everyday life of the setting, which frames the study. Qualitative research positions the researcher as the primary research instrument in a naturalistic environment to collect and interpret participant perspectives (Patton, 2002). In this research, data were collected in the form of detailed observation field notes and in-depth interviews from teacher participants. Analysis of qualitative data was grounded on the ideals that these participants perceived for their world (Corbin & Strauss, 2008). Data were presented in categories and themes to describe lived experiences of participants within the *SHL* curriculum implementation process.

Research Sampling

A selective sample of schools was chosen from the 15 schools randomly assigned to implement the *SHL Curriculum*. Any experimental school ($n = 15$) that had at least one new *SHL* teacher during the 2012-2013 school year was considered for the study. Of the 15 schools with teachers new to the project, four schools located in a cluster were

selected to facilitate data collection. Implementation fidelity in each of the sixth-grade physical education classes (n = 17) at the four schools was examined in this study.

Research Sites and Participants

Participants were middle school physical education teachers and their sixth-grade students from Albany, New River, St. Anthony, and Titan Middle Schools (pseudonyms). These middle schools were located in the Piedmont region of NC and served students in Grades 6-8; ages 11-14. Each school was selected because the teachers were voluntarily participating in the *SHL* project effectiveness study for the first time during the 2012-2013 school year. The four schools selected for this research had been randomly assigned to the intervention group. Physical education teachers taught the *SHL* curriculum consisting of the 20 structured lessons from the sixth-grade *Cardio Fitness Club* unit.

Albany Middle School enrolled over 220 sixth-grade students during the 2012-2013 academic year. Approximately 52% of the students were male, 14% were White, 69% were Black, 7% were Hispanic, 6% were Asian, and 73% received free and reduced-price meals (FARM; Education First NC School Report Cards, 2011). There were four middle school physical education teachers (two males and two females) employed at Albany Middle School during the 2012-2013 academic year. Students received gender segregated physical education instruction for the entire school year on a 1-week physical education and 1-week health education rotation. There were seven sixth-grade physical education classes at Albany Middle School.

New River Middle School enrolled over 190 sixth-grade students during the 2012-2013 academic year. Approximately 50% of the students were male, 35% were White, 12% were Black, 48% were Hispanic, and 77% received FARM (Education First NC School Report Cards, 2011). There were two middle school physical education teachers (one male and one female) employed at New River Middle School during the 2012-2013 academic year. Students received coeducational physical education instruction for one semester using a 4-week physical education and 2-week health education rotation. Only the sixth-grade students enrolled in three physical education classes during the first semester participated in this study.

St. Anthony Middle School enrolled over 180 sixth-grade students during the 2012-2013 academic year. Approximately 48% of the students were male, 52% were White, 17% were Black, 25% were Hispanic, and 60% received FARM (Education First NC School Report Cards, 2011). There were two middle school physical education teachers (one male and one female) employed at St. Anthony Middle School during the 2012-2013 academic year. Students received coeducational physical education instruction for one semester using a 4-week physical education and 2-week health education rotation. Only the sixth-grade students enrolled in three physical education classes during the first semester participated in this study.

Titan Middle School enrolled over 200 sixth-grade students during the 2012-2013 academic year. Approximately 55% of the students were male, 24% were White, 44% were Black, 25% were Hispanic, and 86% received FARM (Education First NC School Report Cards, 2011). There were two middle school physical education teachers (one

male and one female) employed at Titan Middle school during the 2012-2013 academic year. Students received gender segregated physical education instruction on an A day/B day rotation. There were four sixth-grade physical education classes at Titan Middle School.

Variables and Measures

Adherence – Coverage and Delivery Quality (RQ1)

Lesson observation was the primary data source for adherence data. As such it was the primary FOI variable in this study.

Lesson observations. I collected detailed field notes for each observed lesson and systematically coded these to examine adherence based on coverage and delivery quality per sixth-grade physical education class.

Intervention Outcomes (RQ2)

Standardized knowledge tests. Student knowledge test gain scores served as one of two intervention outcomes in this study. All sixth-grade students receiving the intervention completed a 10-item pre- and post-instruction knowledge test. The two assessments were multiple-choice (four choices) and administered through Qualtrics (Qualtrics Labs, Inc., Provo, UT). The test items for the *Cardio Fitness Club* unit were created to match the content covered in the unit to increase the content validity of the tests. The pretest was administered during the lesson immediately preceding lesson 1 and the posttest was administered immediately after lesson 20.

In-class student physical activity intensity levels. The second outcome variable, student physical activity intensity levels, was measured with Actigraph GT3X+

accelerometers (Actigraph, Pensacola, FL). The GT3X+ accelerometer is a small (4.6 cm × 3.3 cm × 1.5 cm), light (19 g) device contained in a plastic case that participants attach using an adjustable elastic waist strap. The solid state accelerometer measures and records physical activity vector magnitude counts in three physical axes. In the pre-filtered raw mode, accelerometer output represents the actual G-force signal untransformed by a band-pass filter and sampled at a frequency of 30 Hz (John & Freedson, 2012). The G-force signal is then converted by a 12-bit analog-to-digital converter with the output band-pass filtered across a frequency range of 0.25 Hz to 2.5 Hz (John & Freedson, 2012). The filtered signal is then rectified and integrated over a user-specified time frame (i.e., epoch). At the end of each epoch, a vector magnitude count is calculated and stored in memory, the integrator is reset, and the process starts over. The sampling epoch length was set at 10 s for this study.

Intensity levels were categorized as sedentary, light, moderate, vigorous, and very vigorous. Intensity level category cutoff points (sedentary: vector magnitude count < 150; light: 150-499; moderate: 500-3,999; vigorous: 4,000-7,599; very vigorous: > 7,599) categorized by Freedson, Pober, and Janz (2005) determined student physical activity intensity levels (i.e., sedentary, light, moderate, vigorous, or very vigorous) during *SHL* lessons. The device has demonstrated high inter-instrument reliability (r = .86-.89) and intra-instrument reliability (r = .86; Melanson & Freedson, 1995).

Strategies to Facilitate Implementation (RQ2)

Professional development attendance. Teacher professional development was a critical component of the *SHL* project and as such was a potential moderator of

implementation fidelity. Attendance sheets from *SHL* professional development sessions were used to determine how many and which professional development sessions each participating teacher attended during the study.

Coaching visits. The support provided by the *SHL* project team also was a critical component of the *SHL* project. The level of support provided by the trained project coaches to the project teachers can be a moderating factor on teacher fidelity. Coach's travel logs were used to determine the number and purpose of coaching visits per sixth-grade physical education class.

Participant Responsiveness (RQ2)

Ease of Use survey. In this study, participant responsiveness was the teachers' satisfaction level with the intervention components. An "Ease of Use Survey" was completed by the teachers in this study at the completion of the *Cardio Fitness Club* unit. The survey was web-based and disseminated through Qualtrics. The survey contained items pertaining to the quality of the lessons, student reflective journal, coaching, professional development, and teacher resources.

Context (RQ3)

Teacher interviews. In this study, preexisting and intervention context factors were defined through the reasons teachers gave for modifying the intervention lessons. I formulated interview questions during the lesson observations as teachers were observed adhering and deviating from the intervention lessons. The goal of the teacher interviews was to gain insight into why teachers made changes to the intervention lessons and the

teachers' positive and negative assessments about the appropriateness of the intervention at their school.

Procedures

The procedures proposed for this dissertation study consisted of three phases. Table 3.1 provides the timeline of the procedures. In Phase 1, institutional review board (IRB) approval was obtained from the University of North Carolina at Greensboro. Because this research was conducted within an on-going research study with IRB approval (11-0009), the principal investigator reported this research as a "research modification." The IRB approval letter is provided in Appendix A. Phase 2 entailed 10-15 weeks of data collection. Phase 3 included data analysis and the communication of the findings. More detailed procedures can be found in the following paragraphs.

Human Participants and Ethics

Upon dissertation proposal approval, I submitted an IRB application to the principal investigator outlining my research. The principal investigator then submitted a modification to the current IRB application (11-0009) to the University of North Carolina at Greensboro's Office of Research Compliance and Guilford County Schools. Teachers participating in this study provided informed consent prior to data collection (see Appendix B). Students selected to wear accelerometers signed an assent form (see Appendix B) and their parent and/or guardian provided parental permission prior to collecting physical activity intensity levels data (see Appendix B). I filed teacher consent forms, assent forms, and parental permission forms in a locked file cabinet in a locked

research laboratory at the University of North Carolina at Greensboro. All schools participating in this study were assigned a pseudonym to maintain their anonymity.

Table 3.1

Tentative Timeline for Dissertation Progression

Phase	Week	Task
1	1-2	IRB approval of 11-0009 modification
2	3-15	Data collection
	3	Obtained signed consent and assent forms and class lists
	4-14	Observed lessons, collected physical activity data, and interviewed teachers
	15	Administered posttests and ease of use surveys
3	16-30	Data reduction, analysis, and dissertation writing
	16-17	Reduced physical activity, knowledge test, and Ease of Use survey data
	18-19	Reduced observation and interview data
	20-22	Analyzed data
	23-30	Wrote chapters IV, V, and VI of dissertation

All teachers and students participating in this dissertation study were assigned a unique identification number to maintain their anonymity. All electronic files containing school names and their pseudonym and teachers' and students' names and their assigned identification numbers were stored in a separate folder in a password protected computer

in the electronically keyed research laboratory at the University of North Carolina at Greensboro. All schools and teacher and student data collected were identified using the appropriate identifier.

Data Collection

I informed participants of the purpose and processes to be used in the research and secured assent, consent, and permission from the participants as described above. I asked each participating teacher for his/her sixth-grade physical education class enrollment lists for the 2012-2013 school year. Class lists were used to assign students unique study identification numbers.

Lesson observations (RQ1). Once knowledge pretesting was complete at each participating school, participating teachers began teaching the sixth-grade *Cardio Fitness Club* unit (lessons 1-20) to their sixth-grade students. I began visiting each participating school beginning with lesson 6 to collect detailed field notes on lessons 6-20. Observation field notes for lessons 1-5 were not collected to allow teachers to become familiar with the lesson format, pacing, and content prior to data collection. Observation field notes were recorded on the standard qualitatively-oriented *SHL* observation form. The *SHL* observation form was used because it aligned with the general structure of each lesson and facilitated field notes collection. It contained sections for recording the standard school, teacher, grade, period, and date of observation and categories for describing in detail events that occurred during each of the 5 Es within the lessons. The categories for the 5 Es contained subcategories related to lesson timing, delivery quality, student enjoyment, and student journal use during that section. I observed, in a non-

participatory capacity, at least 40% of lessons 6-20 implemented in each sixth-grade class. Participating teachers were not informed about when or which lessons I would observe.

Standardized knowledge growth (RQ2). Depending on school academic calendars and school computer lab availability, participating teachers administered the sixth-grade, 10-item standardized knowledge pretest to their sixth-grade physical education students prior to teaching *Cardio Fitness Club* lesson 1. I sent a Qualtrics pretest hyperlink to teachers via email communication. Teachers administered the pretest to their students in the schools' computer lab or via individual student laptop computers or tablets. After completing the 20 lessons, participating teachers re-administered the 10-item test used for pretesting via Qualtrics. Class completion rates were tracked in Qualtrics during both the pre- and posttests and teachers were reminded of classes that had not completed the testing.

In-class student physical activity intensity levels (RQ2). After pretesting was completed at each school, each participating teacher assisted me in selecting six students per sixth-grade class (n = 102 students). Students were selected purposefully to generate a balanced sample based on gender and body mass index to wear physical activity accelerometers.

During lessons 6-20 of the sixth-grade *Cardio Fitness Club* unit, I collected physical activity intensity level data each time I observed a lesson. The six students selected per class were introduced to and assigned a uniquely numbered accelerometer. Each student wore their assigned accelerometer at least once during lessons 1-5 to learn

how to put it on and to minimize the reactivity effect of wearing the accelerometer for the first time. Students put on the accelerometer as they entered the gymnasium from the locker room and returned it just prior to reentering the locker room at the end of the class period. I observed each student attach the accelerometer to his/her waist to ensure the proper positioning of the accelerometer. Students were asked to wear the device on their waist above their right knee using a supplied elastic waist band. During physical activity intensity level data collection, I recorded lesson unit and number and lesson start and end time.

Professional development attendance (RQ2). During the *Cardio Fitness Club* unit, participating teachers were invited to attend three 6-hr professional development sessions hosted by the larger *SHL* project. The professional development sessions further explained the intervention theory, presented demonstration lessons containing best practices, and provided an open forum for participating teachers to discuss their project experiences. Teachers attending the professional development sessions were asked to sign an attendance sheet. Frequency of attendance (n = 0-3) was the variable of interest in this category.

Coaching visits (RQ2). As a participant in the *SHL* project, each school was assigned a trained project coach. The coach's responsibilities were to provide support to and serve as the first contact for the schools. I was assigned as the coach for the four participating schools in this study. Each time I visited a school and a sixth-grade physical education class to provide project support, I recorded the trip on a travel log.

Ease of Use survey (RQ2). At the completion of the sixth-grade *Cardio Fitness Club* unit, I sent participating teachers a Qualtrics hyperlink via email communication directing them to *a Cardio Fitness Club* unit Ease of Use survey. Teachers were asked to complete the survey within 5 days of finishing lesson 20 of the *Cardio Fitness Club* unit.

Interviews (RQ3). After each lesson observation, I interviewed each teacher asking unique questions based on the recently completed lesson observation. The questions were based on observed deviations from the lessons as written and probed teachers on why they elected to make changes to the intended lessons.

Data Reduction

Data from lesson observations and teacher interviews were transcribed into separate Microsoft Word files. Standardized knowledge test data, physical activity data, Ease of Use survey responses, coaching visits, and professional development attendance were entered into a Statistical Package for the Social Sciences (IBM SPSS Statistics Version 20, Release 20.0.0, IBM, Inc., Armonk, NY) database. Quantitative data sources were examined carefully to check for incomplete or missing data. Qualitative data sources were transcribed for analysis.

Lesson observations and interviews (RQ1 and RQ3). I transcribed the lesson observation field notes and teacher interview responses into individual Microsoft Word files organized by school and teacher observation and interview dates. The transcription process allowed me to become more acquainted with the data.

Standardized knowledge growth (RQ2). Student standardized knowledge test item responses were downloaded from Qualtrics to SPSS and organized by teacher and

class. On both the pre- and posttests, individual item responses were converted from answer selected (i.e., 1 = A, 2 = B, 3 = C, and 4 = D) to answer correct (coded as one) or incorrect (coded as zero). Pre- and posttest total raw scores were determined for each student by summing their individual item responses. Each student's pretest raw score was then subtracted from his/her corresponding posttest raw score to produce their change score. The students' change scores were averaged in each class to produce a class change score average.

In-class student physical activity intensity levels (RQ2). In-class student physical activity intensity levels were recorded as vector magnitudes. Vector magnitudes were downloaded from ActiLife 6 to SPSS. Total vector magnitude for each measured student was converted to vector magnitude per minute to reflect in-class physical activity intensity level. Vector magnitude per minute was determined by dividing the accelerometer physical activity count sum by the total minutes (lesson end time minus lesson start time) in the respective *SHL* lesson. Each student's in-class physical activity intensity level per observed lesson per class was averaged to produce a class physical activity intensity level.

Professional development attendance (RQ2). The number of times each teacher attended a *SHL* professional development session (n = 0-3) was entered into SPSS.

Coaching visits (RQ2). The number of coaching visits was transferred from the travel logs to SPSS organized by sixth-grade physical education class/teacher.

Each sixth-grade physical education class's coaching visit score was added to the professional development score belonging to the teacher of that particular sixth-grade physical education class.

Ease of use survey (RQ2). I downloaded Ease of Use survey data from Qualtrics to SPSS organized by teacher. Survey responses were summed to produce a single survey score per teacher.

Data Analysis

The reduced data were subject to the following data analyses. Data were analyzed to answer the three research questions. Inductive coding techniques were employed to answer Research Questions 1 and 3. Multiple regression was used to answer Research Question 2.

RQ1. *To what extent do teachers implement lessons from a sixth-grade physical education curricular intervention with fidelity (adherence – coverage and delivery quality)?*

Inductive analysis was used to analyze the transcribed lesson observations immediately following each school visit. "Inductive analysis involves discovering patterns, themes, and categories in one's data. Findings emerge out of the data, through the analyst's interactions with the data …" (Patton, 2002, p. 453). Utilizing the constant comparative method (Glaser, 1978), I looked for codes and categories in the data. As described by Corbin and Strauss (2008), I first employed open coding of the data to determine whether there were recurring topics that could be classified into categories and labeled. I elected to open code the observation field notes by hand to facilitate constant

comparison and become more acquainted with the data. The generated open codes described and referred to actual events in the physical education classes. Open coding allowed me to divide the data into distinct categories, look for similarities and differences within the observation field notes, and ask questions about phenomena revealed in the data (Corbin and Strauss, 2008). The open coding procedure allowed me to concentrate on the events observed in the physical education classes and guided my observations as I explored cases of teacher fidelity.

Once categories were developed and labeled, I looked for relationships among the categories using axial coding. During the axial coding process, I developed subcategories focusing on phenomena, antecedents, contexts, and any intervening conditions. As suggested by Corbin and Strauss (2008), I ended my analysis with axial coding because I only was interested in using some of the tools of grounded theory and was not focused on developing theory to answer Research Question 1.

Using the open and axial codes (see Appendix C), I developed a rubric that represented the patterns reflective of the teachers' adherence, delivery quality, and coverage. I used the rubric to quantify the teachers' FOI per observed lesson.

RQ2. *What is the relationship between implementation fidelity and student outcomes (i.e., knowledge gain and physical activity intensity levels)?*

Multiple regression analysis was utilized to answer Research Question 2. Descriptive and correlational analyses were conducted to discover the significance of the predictor variables in contributing to the criterion variable. Unstandardized coefficient beta coefficients and standardized beta coefficients for each predictor variable were

examined to check the statistical significance and relative importance of each predictor variable. Additionally, R^2 and adjusted R^2 were used to examine the relationships between the various predictor variables and the criterion variable.

Regression diagnostic statistics were calculated with the multiple regression analyses to check data for cases that exerted undue influence over the parameters of the model and/or were extreme outliers (Field, 2009). Deviant cases, either individually or in combination with other cases, can significantly influence regression statistics. The following regression diagnostics were examined to check for influential data points: (a) Cook's distance (a measure of the overall influence of a case on the model), (b) leverage or hat values (gauges the influence of the observed value of the outcome variable over the predicted values), and (c) standardized DFBETA (detects cases that influence the regression coefficient). Additionally, collinearity diagnostics (i.e., tolerance and variance inflation factor) were conducted to ensure that no variables were closely related to one another.

RQ3. *What reasons do teachers offer for making changes to the intended curriculum as they implement it in their school contexts (moderating variable of context)?*

Similar to Research Question 1, open and axial coding was used to answer Research Question 3. Codes and categories were developed from the interview transcripts using the constant comparative method (Glaser & Strauss, 1967) whereby line, sentence, and paragraph segments of the interviews were reviewed to decide what codes fit the concepts suggested by the data.

In addition to open and axial coding, selective coding was employed. During selective coding, categorical relationships defined during axial coding were connected and further refined to answer the research question. Central themes were related to each category through the use of explanatory relationship statements and then named and connected to the existing literature.

Threats to Validity, Reliability, and Credibility

Internal Validity

There were several threats to the internal validity of this research. First, participants (i.e., teachers implementing the intervention) were not randomly selected. The schools at which the teachers work, however, were randomly assigned to the treatment group to minimize the selection threat to internal validity. Additionally, the teachers' supervisors (i.e., school administrators and curriculum facilitators) expected the teachers to implement the intervention and, as a result, helped minimize the maturation threat to internal validity. Second, teachers were interviewed at the completion of lessons. Interviews can create environments to fabricate the truth. Interviewees may want to provide the answers that the interviewers want to hear as opposed to providing their honest answers. This threat was minimized by asking direct, narrowly focused questions directing the interviewees to a particular event they created in their teaching and answers were triangulated with class observation/field note data. Finally, the student knowledge test, itself, could have threatened internal validity. Knowledge tests can be biased and not measure the intended variable(s) if they are poorly constructed. This

threat was minimized because test developers designed the knowledge test questions to directly align with the content covered in the *Cardio Fitness Club* unit.

External Validity

The generalizability of the study results depended on the sample representativeness of the population. The participants were sixth-grade physical education teachers (and their students) recruited to implement the *SHL* curriculum. The teachers worked at schools that were recruited and sampled as a part of the larger *SHL* project to represent the middle schools in the Piedmont region and state of North Carolina.

Another threat to external validity was using poorly defined variables. All variables were defined operationally to be meaningful in settings beyond the current study to minimize this threat to external validity. My role as the researcher also was a threat to external validity. Teachers were aware that I was observing and, as a result, could have reacted differently to the intervention, taking greater care to implement the curriculum faithfully; compared to participants who were unaware they were being observed. I informed the teachers that I was observing the curriculum and not their faithfulness to the curriculum.

Reliability

Student physical activity intensity levels were measured with Actigraph GT3X+ accelerometers. The Actigraph GT3X+ device previously demonstrated high inter-instrument reliability (r = .86-.89) and intra-instrument reliability (r = .86; Melanson & Freedson, 1995).

Credibility

Five primary techniques were utilized to maximize the credibility of the information gathered in this study. First, I spent 5 weeks interacting with the participants and observing their sixth-grade physical education classes. This time frame allowed me to gain the trust of the participants and allowed the participants to become comfortable with my presence in their classes. Second, I triangulated the data sources (i.e., lessons as written, detailed observation field notes, and interviews) to enhance the believability of the study (Patton, 2002). Third, I used member checks providing teachers with copies of their transcribed interviews and asking them to confirm and provide feedback regarding the accuracy of the transcriptions and any identified categories and themes (Creswell, 2011). Fourth, I wrote rich, thick descriptions of classroom observations to convey the nature of the class context and the extent to which teachers followed the lesson plans provided by the intervention developers (Creswell, 2011). Fifth, I utilized a competent peer debriefer throughout data collection and analysis to maintain research ethics and maximize the trustworthiness of the data (Lincoln & Guba, 1985).

Role of the Researcher

I believe physical education is a valuable and necessary school subject for current and future K-12 students. As evidence continues to mount suggesting that improvements in fitness help protect against obesity, certain cancers, cardiovascular disease, and in general, premature death, I support the inclusion of fitness activities in physical education. I believe, however, that physical education as physical activity alone does not teach students the skills, knowledge, and abilities to be physically active for a lifetime.

Physical education should include opportunities for students to participate in cognitively engaging physical activities so they can learn the skills necessary not only to be healthy and fit today, but also in their future adult lives. I believe quality physical education programs should appeal to and engage every student regardless of their fitness levels, abilities, or interests. Students' progress should be evaluated using both informative and formative assessments to provide students opportunities to give and receive feedback about their physical and cognitive development.

I was a member of the *SHL* project research team and took a deep interest in the intervention fidelity of the *SHL* curriculum. I served as a research assistant in the project for over a year. In this role, I disseminated materials to the teachers and collected and analyzed project data. I also worked with teachers one-on-one to promote the best environment possible to teach the intervention lessons and attended and participated in weekly project team meetings. In this process I developed a good rapport with teachers and worked to support their efforts to teach the curriculum. Further, I relayed their concerns to the project leaders and advocated for their concerns when feasible.

I am a supporter of the *SHL* curriculum and believe the faithful implementation of its lessons will increase the knowledge and physical activity levels of the student participants. I believe the *SHL* curriculum will provide physical education teachers with a theoretically-based curriculum that adds value to their physical education programs. During my time working with the *SHL* project, I gained a thorough understanding of the curriculum. Additionally, I built relationships with participating teachers and came to know their teaching styles and philosophies of education. Although, I may have brought

some biases into the study, I do not believe they negatively impacted the analysis or interpretation of the findings. To minimize these possible biases, I removed teacher names from data, included measures of triangulation, used multiple data sources, and used a debriefer to review the collected data.

CHAPTER IV

RELATIONSHIP BETWEEN TEACHER FIDELITY AND PHYSICAL EDUCATION STUDENT OUTCOMES

Abstract

As educational practices continue to receive strong criticisms, researchers increasingly concentrate on the influence of scientifically-based curricula and teaching methods on student outcomes. To evaluate the influence of scientifically-based curricula intended for the physical education classroom, researchers should consider fidelity or the extent to which implementation adheres to the model of curriculum and instruction inherent in the research design. The purpose of this study was to assess physical education teachers' fidelity of implementation as they implemented a research-based physical education curricular intervention and examine the relationship between teachers' fidelity to the curricular intervention and student outcomes (i.e., knowledge growth and physical activity intensity levels). Six teachers from 3 middle schools taught 1, 20-lesson unit from a physical education intervention to their 6th-grade students. Teachers' fidelity of implementation was documented using non-participatory observations. Students' knowledge acquisition was measured on a standardized knowledge test prior to and after the 20-lesson unit. Students' physical activity intensity levels were measured on lesson observation days using accelerometers. Open and axial coding was used to develop a rubric to quantify the teachers' fidelity of implementation scores. Multiple regression

with fidelity score as the predictor and knowledge acquisition as the criterion variable, indicated that teachers' fidelity scores accounted for a large portion of variance in student knowledge growth ($R^2 = .79$, adj $R^2 = .74$, $p < .05$). Approximately 48% of all observed lesson time involved moderate to vigorous physical activities. Based on the findings from this research, it appears the more faithful teachers are to teaching research-based curricula as designed, the greater the impact the curricula can have on student achievement.

Keywords: fidelity of implementation, school-based physical education intervention, multiple regression

Introduction

During a time of student, teacher, and school accountability, state legislators are taking control of schools through legislative actions across the United States. Educational mandates and policies related to school vouchers, school performance pay, teacher licensure, and testing legislation have been popular talking points among many legislators despite resistance and criticism from members of the education community. These reforms, along with legislation directed toward specific content areas such as physical education, have redefined the way teachers teach and students learn.

As educational practices continue to receive strong criticisms, researchers increasingly concentrate on the influence of scientifically-based curricula and teaching methods on student outcomes. The need to attend to research-based practices also extends to the field of physical education. Research should be directed toward understanding and measuring classroom environments in which physical education

students move and learn, as well as the educational practices that best respond to the learning characteristics of these students (Sallis et al., 2012).

To evaluate the influence of scientifically-based curricula intended for the physical education classroom, researchers should consider fidelity or the extent to which implementation adheres to the model of curriculum and instruction inherent in the research design (Lillehoj, Griffin, & Spoth, 2004; Lynch & O'Donnell, 2005). Logically, if teachers do not implement research-based curricular interventions as designed, measured outcomes may not be attributable to the effectiveness of the intervention. However, few studies measure fidelity of implementation (FOI) or "the extent to which delivery of an intervention adheres to the protocol or program model originally developed" (Mowbray, Holter, Teague, & Bybee, 2003, p. 315).

The purpose of this study was two-fold. First, I assessed physical education teachers' FOI as they implemented a research-based physical education curricular intervention. Second, I examined the relationship between teachers' fidelity to the curricular intervention and student outcomes. This research was guided by two research questions: (a) to what extent do teachers implement lessons from a sixth-grade physical education curricular intervention with fidelity? and (b) what is the relationship between FOI and student outcomes (i.e., knowledge gain and physical activity intensity levels)?

Assessing FOI increases researchers' and practitioners' understanding of the fit between the intervention and the context. However, the complexity of the implementation context can impact the efficiency and generalizability of the innovation to other situations. Researchers have rarely used fidelity variables to explain intervention

outcomes. Appropriate and accurate fidelity data can inform research and practice to better explain intervention outcomes.

During a time of teacher and student accountability, high stakes testing, and No Child Left Behind mandates, schools and students must have effective curricula to increase achievement outcomes. Furthermore, given the calls by public health organizations to increase physical activity and decrease overweight and obesity levels in children and adolescents, school-based physical education programs need evidence-based interventions proven to increase educational achievement and /or physical activity outcomes. Measuring FOI provides one method of assessing moderators of curricular effectiveness. According to O'Donnell (2008) there is need to conduct research investigating FOI and to understand why and how teachers impact intervention implementation. Studying FOI as part of curricular effectiveness research is critical to understanding intervention success or failure. Valid measures of FOI also can enhance statistical power and explain variance in outcome studies. This research examined variables that enhanced or constrained the effectiveness of a large-scale physical education intervention and the influence of FOI on student outcomes.

Fidelity of Implementation

Educators are increasingly expected to use an assortment of instructional and curricular approaches to address the diverse needs of students and improve student achievement. The No Child Left Behind Act of 2001 (NCLB; 2003) states that educators must use scientifically based teaching methods and strategies proven to be effective. Therefore, researchers and educators should be concerned about the degree to which the

educational interventions they design, implement, and adopt are evidence-based (Lynch & O'Donnell, 2005; Mowbray et al., 2003; Slavin, 2003).

To determine the effectiveness and impact of a research-based curricular intervention it is important to assess the degree to which its implementation adheres to the model of curriculum and instruction inherent in the research design (O'Donnell, 2008). Although there are many competing definitions, I broadly define FOI "as the extent to which delivery of an intervention adheres to the protocol or program model originally developed" (Mowbray et al., 2003, p. 315). Assessing FOI in the context of any intervention study is essential because "failure to establish fidelity can severely limit the conclusions that can be drawn from any outcome evaluation" (Dumas, Lynch, Laughlin, Smith, & Prinz, 2001, p. 39).

In studies assessing FOI, empirical evidence suggests significant correlations between the extent to which intervention studies are implemented with fidelity and the level of treatment outcomes (Blakely, Mayer, Gottschalk, Schmitt, & Davidson, 1987; Dane & Schneider, 1998; Ruiz-Primo, 2006). Although most of these studies have taken place in the context of public health and counseling research (O'Donnell, 2008), the number of FOI studies evaluating the effectiveness of K-12 curricular interventions is increasing (Lillehoj et al., 2004; Lynch & O'Donnell, 2005; O'Donnell, 2008). When evaluating teacher fidelity in the context of such research-based curricula, a five-component FOI framework has been proposed to include adherence, coverage, program differentiation, delivery quality, and participant responsiveness (Lynch & O'Donnell, 2005; O'Donnell, 2008). Given the "... compelling need to better understand how

curricula, instruction, and student diversity affect student achievement in K-12 classrooms" (Lynch & O'Donnell, 2005, p. 2), fidelity studies investigating K-12 curricular intervention effectiveness deserve consideration and have been recommended (O'Donnell, 2008).

Notably, to assess FOI, it is not necessary to evaluate all five components. Researchers should attend to those components that are of interest to their study (Lillehoj et al., 2004; Lynch & O'Donnell, 2005), and few studies report measuring all five. For example, Lynch and O'Donnell (2005) monitored adherence, delivery quality, and participant responsiveness in evaluating teacher fidelity during the implementation of a research-based middle school science curriculum. Similarly, Lillehoj et al. (2004) focused on adherence and delivery quality while assessing fidelity in the context of a middle school problem behaviors study. For purposes of this study, the fidelity components of adherence, coverage, and delivery quality were of primary interest.

Science of Healthful Living Project

I implemented this study as a part of and during the second year of the larger, 5-year *Science of Healthful Living* (*SHL*) project conducted by the University of North Carolina at Greensboro with support provided by the National Institutes of Health. The goal of the *SHL* project was to design and evaluate a science-based middle school *SHL* curriculum to increase students' cognitive knowledge and interest in health-related science, increase their intention to pursue a life science-related career, and improve their family members' understanding of National Institutes of Health funded clinical and basic research results. The *SHL* project designed a health-related science-based curriculum for

middle school physical education. Consisting of two, 20-lesson units, *Cardio Fitness Club* and *Healthy Lifestyles*, the lessons engage students in the scientific inquiry process in physical education. Students engage in physically active lessons to examine the effects of exercise and healthy nutrition on their bodies. They explore a range of topics such as the physiology of exercise, food-fueled energy systems, caloric balance, nutrition/portion sizes, and the role of physical activity and nutrition in stress management. The scientific inquiry process is embedded in each lesson using the 5 Es (i.e., Engagement, Exploration, Explanation, Elaboration, and Evaluation) learning cycle strategy (Bybee et al., 1989) to assist physical educators and students engage in the scientific inquiry process in a hands-on approach to science. Students are afforded opportunities to make predictions about physical activity concepts, test their predictions, and draw conclusions about their predictions in physically active environments. Additional information about the larger *SHL* project can be found at http://www.uncg.edu/hhs/science-of-healthful-living.

During Year 2, the *SHL* project was set in seven local education agencies in the Piedmont region of North Carolina. Twenty-five middle schools, including five new middle schools recruited in Year 2 of the study, participated in the project, including over 16,000 diverse students and 70 middle school physical education teachers. Using a randomized controlled experimental design, the middle schools were stratified during Year 1 into matched pairs according to socioeconomic status (derived from free and reduced-price meals school-level data) and scores on end of grade tests in mathematics and reading. One school in each pair was randomly assigned to teach the *SHL* curriculum

(experimental school) and one school taught the traditional sport-based or multi-activity physical education (control school).

Method

Research Design

I used a mixed methods design to collect and analyze both quantitative and qualitative data examining FOI in the *SHL* project (Creswell, 2013; Creswell & Plano Clark, 2011; Tashakkori & Teddlie, 2010). Using multiple data collection strategies provided a more complete analysis than any single method alone could accomplish when addressing the problem (Creswell, 2013; Creswell & Plano Clark, 2011; Tashakkori & Teddlie, 1998; 2010). A mixed methods design was appropriate because FOI measurement and analysis required both quantitative and qualitative data collection and analysis during various study stages. Mixed methods approaches offered complementary insights into understanding the faithful implementation of the *SHL* curriculum and the impact FOI had on student outcomes.

Research Sampling

A selective sample of schools was chosen from the 15 schools randomly assigned to implement the *SHL Curriculum*. Any experimental school ($n = 15$) that had at least one new *SHL* teacher during the 2012-2013 school year was considered for the study. Of the 15 schools with teachers new to the project, three schools located in a cluster were selected to facilitate data collection. FOI in each participating sixth-grade physical education class ($n = 6$) at the three schools was examined in this study.

Research Sites and Participants

Participants were middle school physical education teachers and their sixth-grade students from Albany, New River, and St. Anthony Middle Schools (pseudonyms). These middle schools were located in the Piedmont region of NC and served students in Grades 6-8; ages 11-14. Each school was selected because the teachers were voluntarily participating in the *SHL* project effectiveness study for the first time during the 2012-2013 school year. The three schools selected for this research had been randomly assigned to the intervention group. Physical education teachers taught the *SHL* curriculum consisting of the 20 structured lessons from the sixth-grade *Cardio Fitness Club* unit.

Albany Middle School enrolled over 220 sixth-grade students during the 2012-2013 academic year. Approximately 52% of the students were male, 14% were White, 69% were Black, 7% were Hispanic, 6% were Asian, and 73% received free and reduced-price meals (FARM; Education First NC School Report Cards, 2011). There were three middle school physical education teachers (one male and two females) employed at Albany Middle School during the 2012-2013 academic year. Students received gender segregated physical education instruction for the entire school year on a 1-week physical education and 1-week health education rotation. The two female teachers and their three sixth-grade classes participated in this study. The two sixth-grade physical education classes assigned to the same teacher were combined in this study to form one sixth-grade physical education class.

New River Middle School enrolled over 190 sixth-grade students during the 2012-2013 academic year. Approximately 50% of the students were male, 35% were White, 12% were Black, 48% were Hispanic, and 77% received FARM (Education First NC School Report Cards, 2011). There were two middle school physical education teachers (one male and one female) employed at New River Middle School during the 2012-2013 academic year. Students received coeducational physical education instruction for one semester using a 4-week physical education and 2-week health education rotation. Only the sixth-grade students enrolled in three physical education classes during the first semester participated in this study. Two of the three physical education classes occurred at the same time and place and were team taught. These two classes were combined to form one class and assigned to the lead teacher.

St. Anthony Middle School enrolled over 180 sixth-grade students during the 2012-2013 academic year. Approximately 48% of the students were male, 52% were White, 17% were Black, 25% were Hispanic, and 60% received FARM (Education First NC School Report Cards, 2011). There were two middle school physical education teachers (one male and one female) employed at St. Anthony Middle School during the 2012-2013 academic year. Students received coeducational physical education instruction for one semester using a 4-week physical education and 2-week health education rotation. Only the sixth-grade students enrolled in three physical education classes during the first semester participated in this study. Two of the three physical education classes occurred at the same time and place and were team taught. These two classes were combined to form one class and assigned to the lead teacher.

Data Sources

Lesson observations. I collected detailed field notes for each observed lesson and systematically coded the field notes to examine adherence based on coverage and delivery quality per sixth-grade physical education class.

Standardized knowledge tests. Student knowledge test gain scores served as one of two intervention outcomes in this study. All sixth-grade students receiving the intervention completed a 10-item pre- and post-instruction knowledge test. The two assessments were multiple-choice (four choices) and administered through Qualtrics (Qualtrics Labs, Inc., Provo, UT). The test items for the *Cardio Fitness Club* unit were created to match the content covered in the unit to increase the content validity of the tests. The pretests were administered during the lesson immediately preceding lesson 1 and the posttests were administered immediately after lesson 20.

In-class student physical activity intensity levels. The second outcome variable, student physical activity intensity levels, was measured with Actigraph GT3X+ accelerometers (Actigraph, Pensacola, FL). The GT3X+ accelerometer is a small (4.6 cm × 3.3 cm × 1.5 cm), light (19 g) device contained in a plastic case that participants attach using an adjustable elastic waist strap. The solid state accelerometer measures and records physical activity vector magnitude counts in three physical axes. In the pre-filtered raw mode, accelerometer output represents the actual G-force signal untransformed by a band-pass filter and sampled at a frequency of 30 Hz (John & Freedson, 2012). The G-force signal is then converted by a 12-bit analog-to-digital converter with the output band-pass filtered across a frequency range of 0.25 Hz to 2.5

Hz (John & Freedson, 2012). The filtered signal is then rectified and integrated over a user-specified time frame (i.e., epoch). At the end of each epoch, a vector magnitude count is calculated and stored in memory, the integrator is reset, and the process starts over. The sampling epoch length was set at 10 s for this study.

Intensity levels were categorized as sedentary, light, moderate, vigorous, and very vigorous. Intensity level category cutoff points (sedentary: vector magnitude count < 150; light: 150-499; moderate: 500-3,999; vigorous: 4,000-7,599; very vigorous: > 7,599) categorized by Freedson, Pober, and Janz (2005) determined student physical activity intensity levels (i.e., sedentary, light, moderate, vigorous, or very vigorous) during *SHL* lessons. The device has demonstrated high inter-instrument reliability (r = .86-.89) and intra-instrument reliability (r = .86; Melanson & Freedson, 1995).

Data Collection

Lesson observations. Once knowledge pretesting was complete at each participating school, participating teachers began teaching the sixth-grade *Cardio Fitness Club* unit (lessons 1-20) to their sixth-grade students. I began visiting each participating school beginning with lesson 6 to collect detailed field notes on lessons 6-20. Observation field notes for lessons 1-5 were not collected to allow the teachers to become familiar with the lesson format, pacing, and content prior to data collection. Observation field notes were recorded on the standard qualitatively-oriented *SHL* observation form. The *SHL* observation form was used because it aligned with the general structure of each lesson and facilitated field notes collection. It contained sections for recording the standard school, teacher, grade, period, and date of observation and categories for

describing in detail events that occurred during each of the 5 Es within the lessons. The categories for the 5 Es contained subcategories related to lesson timing, delivery quality, student enjoyment, and student journal use during that section. I observed, in a non-participatory capacity, at least 40% of lessons 6-20 implemented in each sixth-grade class. Participating teachers were not informed about when or which lessons I would observe.

Standardized knowledge growth. Depending on school academic calendars and school computer lab availability, participating teachers administered the sixth-grade, 10-item standardized knowledge pretest to their sixth-grade physical education students prior to teaching *Cardio Fitness Club* lesson 1. I sent a Qualtrics pretest hyperlink to teachers via email communication. Teachers administered the pretest to their students in the schools' computer lab or via individual student laptop computers or tablets. After completing the 20 lessons, participating teachers re-administered the 10-item test used for pretesting via Qualtrics. Class completion rates were tracked in Qualtrics during both the pre- and posttests and teachers were reminded of classes that had not completed the testing.

In-class student physical activity intensity levels. After pretesting was completed at each school, each participating teacher assisted me in selecting six students per sixth-grade class (n = 36 students). Students were selected purposefully to generate a balanced sample based on gender and body mass index to wear physical activity accelerometers.

During lessons 6-20 of the sixth-grade *Cardio Fitness Club* unit, I collected physical activity intensity level data each time I observed a lesson. The six students selected per class were introduced to and assigned a uniquely numbered accelerometer. Each student wore their assigned accelerometer at least once during lessons 1-5 to learn how to put it on and to minimize the reactivity effect of wearing the accelerometer for the first time. Students put on the accelerometer as they entered the gymnasium from the locker room and returned it just prior to reentering the locker room at the end of the class period. I observed each student attach the accelerometer to his/her waist to ensure the proper positioning of the accelerometer. Students were asked to wear the device on their waist above their right knee using a supplied elastic waist band. During physical activity intensity level data collection, I recorded lesson unit and number and lesson start and end time.

Data Reduction

Lesson observations. I transcribed the lesson observation field notes into individual Microsoft Word files organized by school and teacher observation dates. The transcription process allowed me to become more acquainted with the data.

Standardized knowledge growth. Student standardized knowledge test item responses were downloaded from Qualtrics to the Statistical Package for the Social Sciences (SPSS; IBM SPSS Statistics Version 20, Release 20.0.0, IBM, Inc., Armonk, NY) and organized by teacher and class. On both the pre- and posttests, individual item responses were converted from answer selected (i.e., 1 = A, 2 = B, 3 = C, and 4 = D) to answer correct (coded as one) or incorrect (coded as zero). Pre- and posttest total raw

scores were determined for each student by summing their individual item responses. Each student's pretest raw score was then subtracted from his/her corresponding posttest raw score to produce their change score. The students' change scores were averaged in each class to produce an average class change score (see Table 4.1).

Table 4.1

Class-Level Student Outcomes

	# of Students	Knowledge Change (*SD*)	PA Intensity Level (*SD*)
Albany Middle			
Teacher 1	23	13.9% (21.3)	2,260.03 CPM (225.75)
Teacher 2[a]	55	12.0% (20.9)	2,069.33 CPM (85.47)
New River Middle			
Teacher 3[b]	45	22.0% (19.4)	2,186.58 CPM (351.16)
Teacher 4	28	15.7% (19.9)	1,734.3 CPM (207.44)
St. Anthony Middle			
Teacher 5[b]	55	17.8% (20.4)	2,520.47 CPM (426.65)
Teacher 6	23	20.9% (19.1)	2,468.28 CPM (117.57)

Note. CPM = vector magnitude count per minute, [a]Outcome data were combined from two physical education classes assigned to this teacher., [b]Teacher team taught and absorbed the students assigned to the other teacher assisting with the lesson.

In-class student physical activity intensity levels. In-class student physical activity intensity levels were recorded as vector magnitudes by the Actigraph GT3X+ accelerometers. Vector magnitudes were downloaded from ActiLife 6 (Actigraph, Pensacola, FL) to SPSS. Total vector magnitude for each measured student was converted to vector magnitude per minute to reflect in-class physical activity intensity

level. Vector magnitude per minute was determined by dividing the accelerometer physical activity count sum by the total minutes (lesson end time minus lesson start time) in the respective *SHL* lesson. Each student's in-class physical activity intensity level per observed lesson per class was averaged to produce an average class physical activity intensity level (see Table 4.1).

Data Analysis

Inductive analysis was used to analyze the transcribed lesson observations immediately following each school visit. "Inductive analysis involves discovering patterns, themes, and categories in one's data. Findings emerge out of the data, through the analyst's interactions with the data …" (Patton, 2002, p. 453). Utilizing the constant comparative method (Glaser, 1978), I looked for codes and categories in the data. As described by Corbin and Strauss (2008), I first employed open coding of the data to determine whether there were recurring topics that could be classified into categories and labeled. I elected to open code the observation field notes by hand to facilitate constant comparison and become more acquainted with the data. The generated open codes described and referred to actual events in the physical education classes. Open coding allowed me to divide the data into distinct categories, look for differences and similarities within the observation field notes, and ask questions about phenomena revealed in the data (Corbin and Strauss, 2008). The open coding procedure allowed me to concentrate on the events observed in the physical education classes and guided my observations as I explored cases of teacher fidelity.

Once categories were developed and labeled, I looked for relationships among the categories using axial coding. During the axial coding process, I developed subcategories focusing on phenomena, antecedents, contexts, and any intervening conditions. As suggested by Corbin and Strauss (2008), I ended my analysis with axial coding because I only was interested in using some of the tools of grounded theory and was not focused on developing theory.

Using the open and axial codes (see Appendix C), I developed a rubric that represented the patterns reflective of the teachers' adherence, delivery quality, and coverage. A 42-item dichotomous rubric was developed from the qualitative analysis to measure teacher adherence, coverage, and delivery quality (see Figure 4.1). I then used the rubric to quantify teachers' FOI per observed lesson.

Once the qualitative data analysis was complete, I used the teachers' FOI scores generated by the rubric to further inform the quantitative analysis. Multiple regression was used to measure the contribution of overall lesson FOI and FOI to each of the five Es to the prediction of class knowledge change and class physical activity intensity levels.

Results

During the coding process four major categories emerged from the data: (a) high adherence and good delivery quality, (b) high adherence but poor delivery quality, (c) low adherence but good delivery quality, and (d) low adherence and poor delivery quality. The four categories emerged throughout and across lesson observations.

Science of Healthful Living (*SHL*) Fidelity of Implementation Observation Rubric (Target Schools)

Date of Observation_____ Raw Score = _____/42 = _____
Date Observation Scored_____
School ID_____
Teacher ID_____
Class Period Start and End Time_____
Grade_____
SHL Module and Lesson_____

Please check Yes (1 point) or No (0 points) for each item.

	NO	YES	COMMENTS
1. Teacher and student materials (task cards, journals, equipment) ready before class			
2. Teacher organized and familiar with lesson			
3. Teacher stated the lesson essential question			
4. Teacher referred to the lesson evaluation rubric			
5. Teacher followed the timing of the lesson (or made lesson longer)			
6. Teacher emphasized the intended science concepts throughout the lesson			
Engagement			
7. Engagement occurred as designed or occurred with appropriately substituted activities			
8. Journal used as intended			
9. Teacher models movement tasks			
10. Teacher corrects immature movement patterns			
11. Teacher checks for student understanding before moving on to the next step			
12. Teacher encourages student participation			
13. Students are engaged in the lesson segment			
14. Students are successful during activities			
Exploration			
15. Exploration occurred as designed or occurred with appropriately substituted activities			
16. Journal used as intended			
17. Teacher models movement tasks			
18. Teacher corrects immature movement patterns			
19. Teacher checks for student understanding before moving on to the next step			
20. Teacher supplies students adequate think time			
21. Teacher encourages student participation			
22. Students are engaged in the lesson segment			
23. Students are successful during activities			
Explanation			
24. Explanation occurred as designed or occurred with appropriately substituted activities			
25. Journal used as intended			
26. Teacher checks for student understanding before moving on to the next step			
27. Teacher supplies students adequate think time			
28. Teacher provides students ample opportunities to talk with their peers			
29. Teacher provides students individual turns to speak			
30. Teacher encourages student participation			
31. Students are engaged in the lesson segment			
Elaboration			
32. Elaboration occurred as designed or occurred with appropriately substituted activities			
33. Journal used as intended			
34. Teacher checks for student understanding before moving on to the next step			
35. Teacher supplies students adequate think time			
36. Teacher provides students ample opportunities to talk with their peers			
37. Teacher provides students individual turns to speak			
38. Teacher encourages student participation			
39. Students are engaged in the lesson segment			
Evaluation			
40. Evaluation occurred as designed or occurred with appropriately substituted activities			
41. Journal used as intended			
42. Physical activity homework assigned			

Figure 4.1. Fidelity of implementation observation rubric.

No one category represented any entire single lesson as teachers fluctuated between the four categories as lessons progressed through their critical components. The developed rubric, therefore, contains six sections related to the critical components of the lessons. Within the six sections are statements to prompt the observer to ascertain the level of teacher delivery quality, coverage, and adherence to the lesson as designed. Table 4.2 displays the participating teachers' fidelity scores determined by the 42-item dichotomous rubric.

Table 4.2

Teachers' Average Fidelity of Implementation Scores

	No. of Obs.	Total	Eng.	Explore.	Explan.	Ela.	Eval.
Albany Middle							
Teacher 1	6	.48	.84	.81	.55	.06	.17
Teacher 2	6	.42	.77	1.0	.38	.1	0
New River Middle							
Teacher 3	7	.83	1.0	.82	.95	.77	.47
Teacher 4	7	.75	.83	1.0	.71	.84	.71
St. Anthony Middle							
Teacher 5	7	.8	.83	.8	1.0	.8	.66
Teacher 6	7	.82	1.0	.87	.91	.95	.74

Note. Obs. = observations, Total = overall fidelity score, Eng. = fidelity score during the Engagement, Explore. = fidelity score during the Exploration, Explan. = fidelity score during the Explanation, Ela. = fidelity score during the Elaboration, Eval. = fidelity score during the Evaluation.

Multiple regression was used to assess the contribution of fidelity to the prediction of class knowledge change and class physical activity intensity levels over the

course of the 20-lesson *Cardio Fitness Club* unit. The outcomes of paired sample *t* tests indicated the changes between each class's average pretest score and average posttest score were statistically significant (see Table 4.3). In addition, the correlation between each class's pretest scores and posttest scores were statistically significant (see Table 4.3).

Table 4.3

Differences and Correlations Between Pre- and Posttest Scores

			Differences			Correlations	
	Pretest (%) *M* (*SD*)	Posttest (%) *M* (*SD*)	*t*	*df*	*p**	*r*	*p***
Albany Middle							
Teacher 1	31.7 (15.3)	45.7 (21.1)	3.14	22	.005	.35	.04
Teacher 2	33.1 (15.7)	45.1 (20.5)	4.27	54	<.001	.36	.01
New River Middle							
Teacher 3	27.8 (14.3)	49.8 (20.3)	7.62	44	<.001	.41	.01
Teacher 4	35.4 (16.2)	51.1 (18.5)	4.18	27	<.001	.35	.04
St. Anthony Middle							
Teacher 5	32.9 (14.9)	50.7 (17.8)	6.47	54	<.001	.23	.03
Teacher 6	30.9 (10.0)	51.7 (19.0)	5.25	22	<.001	.26	.04
All Teachers	31.9 (14.8)	48.8 (19.5)	12.58	228	<.001	.32	<.001

Note. *Significant at the $p < .01$ level, **Significant at the $p < .05$ level.

Regression diagnostic statistics were calculated with the multiple regression analyses to check data for cases that exerted undue influence over the parameters of the model and/or were extreme outliers (Field, 2009). Deviant cases, either individually or in combination with other cases, can significantly influence regression statistics. The

following regression diagnostics were examined to check for influential data points: (a) Cook's distance (a measure of the overall influence of a case on the model), (b) leverage or hat values (gauges the influence of the observed value of the outcome variable over the predicted values), and (c) standardized DFBETA (detects cases that influence the regression coefficient). Additionally, collinearity diagnostics (i.e., tolerance and variance inflation factor) were conducted to ensure that no variables were closely related to one another. The obtained variance inflation factor in all cases but one was less than 11.5. Elaboration fidelity score had a strong collinearity with the other predictors and was excluded from all multiple regression models.

Class-level knowledge change and class-level physical activity intensity level were the criterion variables in the multiple regression analyses. Overall fidelity score was the sole predictor variable in the first analysis for each criterion variable, while the four Es' (i.e., Engagement, Exploration, Explanation, and Evaluation) scores were entered in the second analysis for each criterion variable.

Overall fidelity score accounted for 79% [adj $R^2 = .74$, $F(1, 4) = 15.4$, $p < .05$] of the variance in knowledge change. The overall fidelity beta coefficient was statistically significant (see Table 4.4). When the 4 Es were entered as predictors, Engagement, Exploration, Explanation, and Evaluation contributed to the overall model fit for knowledge change [$R^2 = .99$, adj $R^2 = .99$, $F(4, 1) = 416.99$, $p < .05$]. The beta coefficients for Engagement and Explanation were statistically significant (see Table 4.4).

Table 4.4

Regression Output for Knowledge Change

Fidelity Component	*M*	*SD*	Knowledge Change *B*	*SE B*	β	*t*	*p*
Overall	.68	.19	0.19	0.05	.89	3.92	.02
Engagement	.88	.1	0.23	0.01	.58	17.35	.04
Exploration	.88	.09	0.08	0.02	.2	4.5	.14
Explanation	.75	.25	0.13	0.01	.82	10.04	.04
Evaluation	.46	.31	-0.03	0.01	-.24	-3.87	.16
Elaboration	.59	.4	Excluded from model				

Note. Significant at the $p < .05$ level.

Therefore, physical education students taught by physical education teachers with higher fidelity levels displayed more knowledge growth. The specific Es that significantly contributed to knowledge gain included Engagement and Explanation.

Multiple regression targeting class-level physical activity intensity levels did not produce any significant findings using overall fidelity score or the 4 Es' scores (see Table 4.5). The mean proportion of time spent in moderate to vigorous physical activities and sedentary activities over 40 *Cardio Fitness Club* lesson observations was 47.9% (*SD* = 2.4%) and 41.1% (*SD* = 2.7%), respectively (see Table 4.6). Twenty-one of the 40 (52.5%) lesson observations met the Centers for Disease Control and Prevention's recommendation (U.S. Department of Health and Human Services, 2010) of spending at

least 50% of physical education class time participating in moderate to vigorous physical activities.

Table 4.5

Regression Output for Physical Activity Intensity Levels

			Physical Activity Intensity Levels				
Fidelity Component	*M*	*SD*	*B*	*SE B*	β	*t*	*p*
Overall	.68	.19	514.92	437.69	.51	1.18	.31
Engagement	.88	.1	-434.57	1,463.11	-.23	-0.3	.82
Exploration	.88	.09	-1,537.38	1,966.03	-.77	-0.78	.58
Explanation	.75	.25	-131.28	1,388.38	-.17	-0.1	.94
Evaluation	.46	.31	414.15	863.57	.68	0.48	.72
Elaboration	.59	.4	Excluded from model				

Note. Significant at the $p < .05$ level.

Discussion

A limited number of studies have examined the impact FOI has on student outcomes (O'Donnell, 2008). The reviewed studies suggested that high-fidelity implementation enhanced intervention outcomes. The purpose of this study was to assess physical education teachers' FOI as they implemented a research-based physical education curricular intervention and examine the relationship between teachers' fidelity to the curricular intervention and student outcomes. Several findings deserve discussion.

Table 4.6

Mean Percentage of Lesson Time Spent at Different Physical Activity Intensities

	Sedentary[a] (*SD*)	Light[a] (*SD*)	Moderate[a] (*SD*)	Vigorous[a] (*SD*)	Very Vigorous[a] (*SD*)
Albany Middle					
Teacher 1[b]	37.9 (9.4)	14.2 (3.4)	41.5 (7.2)	4.2 (1.9)	2.1 (2.1)
Teacher 2[b]	42.4 (3.6)	12.5 (0.4)	38.6 (3.6)	5.5 (0.6)	1.0 (0.7)
New River Middle					
Teacher 3[c]	37.9 (7.4)	10.8 (2.2)	41.5 (7.1)	7.8 (3.4)	1.9 (1.0)
Teacher 4[c]	44.5 (16.8)	9.4 (2.7)	39.5 (11.6)	5.4 (4.1)	1.2 (0.7)
St. Anthony Middle					
Teacher 5[c]	41.0 (4.0)	8.9 (3.0)	36.8 (4.8)	9.8 (4.9)	3.5 (2.1)
Teacher 6[c]	42.7 (0.1)	10.2 (1.6)	35.1 (1.6)	9.0 (1.8)	3.0 (0.7)
All Teachers	41.1 (2.7)	11.0 (2.0)	38.8 (2.6)	7.0 (2.3)	2.1 (1.0)

Note. [a]Cut-off points suggested by Freedson, Pober, and Janz (2005), [b]Physical activity intensity level data collected on 6 lessons, [c]Physical activity intensity level data collected on 7 lessons.

First, students receiving the *SHL* curricular intervention made statistically significant improvements in their fitness-based knowledge. The mean gain for the six physical education classes was 16.9% (*SD* = 20.3%). However, after inspecting the mean pretest score (*M* = 31.9%, *SD* = 14.8%) and the mean posttest score (*M* = 48.8%, *SD* = 19.5%), students performed below what most educators would consider proficient at the end of the intervention. The majority of the teachers and students in this study were participating in multi-activity physical education programs before agreeing to participate in this study. The mean pretest score suggested that students had very limited prior knowledge of fitness-based concepts. A large body of findings shows that learning

proceeds primarily from relevant prior knowledge, and only secondarily from the taught information (Dochy, Segers, & Buehl, 1999). The limited relevant prior knowledge of the sixth-grade students in the present investigation could have constrained their knowledge gain. It is unclear how much fitness education the students in this study received prior to enrolling in sixth grade, but it is clear that the North Carolina Essential Standards for Healthful Living for Grades 3-5 includes fitness education as a part of its standard course of study (Public Schools of North Carolina, 2011). Additionally, if the students in this study did receive fitness education in fifth grade, the dose and coverage would be questionable given the current trends in instructional time in elementary school physical education.

Second, multiple regression showed that overall FOI accounted for 79% of the variance in fitness-based knowledge growth in sixth-grade students participating in the curricular intervention. Overall FOI made a statistically significant contribution to the prediction of student knowledge gain. Specific Es (i.e., Engagement and Explanation) had a statistically significant effect on knowledge gain. These findings reinforce earlier research demonstrating that FOI has statistically significant implications on student outcomes (Allinder et al., 2000; Butler-Songer & Gotwals, 2005; Hall & Loucks, 1977; Penuel & Means, 2004; Ysseldyke et al., 2003). Therefore, the more faithful teachers are to teaching research-based curricula as designed, the greater the impact the curricula can have on student achievement. During a time of student and teacher accountability, it is critical that educators use curricula proven to be effective.

Although no statistically significant relationships were found between teacher fidelity levels and student physical activity intensity levels, it is worth noting that approximately 48% of all observed lesson time involved moderate to vigorous physical activities. This proportion is slightly higher compared to other studies (e.g., Chow, McKenzie, & Louie, 2009; McKenzie et al., 2006) reporting physical activity intensity levels during physical education lessons. I also found that in over 52% of the observed lessons, students spent at least 50% of the physical education lesson time participating in moderate to vigorous physical activities. While data like these are limited in other studies, it is clear that approximately 48% of the observed lessons did not meet the Centers for Disease Control and Prevention's recommendation for moderate to vigorous physical activity.

As with most research studies, this investigation had a few limitations deserving discussion. First, I relied on six fidelity checks at one school and seven fidelity checks at the other two schools. A more thorough account of teacher FOI and its impact of student outcomes would have been attained with more frequent fidelity checks (e.g., daily). Second, although the NIH randomized control trial used a representative matched sample with control group, the research reported here examined only a subset of this population. Specifically, the first-year (in the intervention) middle school teachers participating in this study were not demographically representative of the sample of teachers participating in the larger NIH sponsored study. Therefore, the study's findings have limited generalizability. Additional research should be implemented using a larger,

demographically heterogeneous sample size to satisfy the statistical assumptions associated with multiple regression.

The findings have potential to address the challenge of implementing effective approaches in practice. Cook, Landrum, Tankersley, and Kauffman (2003) highlighted that approaches may be rendered counterproductive or ineffective when implemented with inadequate fidelity or if not used with appropriate dosage (amount of treatment). Considering this problem in the context of the current investigation, teachers implementing the *SHL* curriculum (a research-based physical education intervention) with lower FOI experienced smaller increases in knowledge gain compared to teachers implementing the intervention with higher FOI. Failing to consider FOI is a major threat to internal validity. The intervention's role in enhancing or constraining outcomes cannot be fully determined without measuring FOI.

Utilizing procedures integrated into the research design for maximizing internal validity enhances the integrity of study findings. This study's findings highlighted that the extent to which the intervention was taught as prescribed (quality), the extent to which the content was delivered (coverage), and the extent to which the internal structure of the lessons were maintained during delivery (structure) can significantly affect student outcomes. Given that few studies to date have assessed the role of FOI on physical education outcomes, future fidelity research in physical education is justified. Monitoring FOI may modify outcomes of existing and future program effectiveness, efficacy, and evaluation studies of physical education curricula.

Additionally, practicing teachers need continual professional development and support to implement and maintain evidence-based practices in their classrooms. To build and maintain teachers' capacities to deliver effective programs, they need frequent and consistent support for a minimum of one year in the form of instruction, feedback, and motivators (National Institute of Child Health and Human Development, 2000). Further, research (e.g., Gersten & Brengelman, 1996) has suggested that successful teacher training for evidenced-based practices follows when content and approaches are similar to the values, beliefs, and goals of important school staff (i.e., lead teachers and administrators). It also is important that coaches and principals use valid and reliable fidelity measures to facilitate continuing mentoring and classroom observations. Structuring observations to identify crucial teaching behaviors offers a method to examine teacher ability/effectiveness to implement evidence-based practices. In this research study, adherence to lesson delivery as intended (i.e., following the lesson structure and script) appeared to increase the contribution of the physical education curriculum to student knowledge gain. Teachers and administrators may consider incorporating fidelity into existing documents used for mentoring or administrative observations.

In summary, while statistically significant growth was evident in the fitness-based knowledge of sixth-grade physical education students, results varied as physical education teachers taught the *SHL* curriculum with different fidelity levels. High fidelity requires time and effort. In spite of years of teaching experience, it requires numerous professional development opportunities to educate teachers and staff about how to

implement evidenced-based, educational interventions. Increasing fitness-based knowledge at the middle school level is possible when scientifically-based physical education instruction is conveyed with fidelity.

CHAPTER V

AN EXAMINATION OF TEACHERS' RATIONALES FOR CHANGING A PHYSICAL EDUCATION INTERVENTION

Abstract

New school-based curricular interventions are fundamental in meeting the diverse needs of today's students and improving student outcomes. In terms of curricular interventions, teachers are the primary implementation agents; therefore, the fidelity of such interventions depends on teacher adoption and delivery. To understand and meet such fidelity challenges, this qualitative research study examined the contextual components and implementation mechanisms contributing to the variation in teachers' perspectives and methods of implementing the *Science of Healthful Living* curriculum; a research-based intervention designed to increase middle school students' fitness-based knowledge through physically active lessons. The results of this study suggested a multitude of preexisting contextual factors, such as lack of instructional time, space, and equipment, influence teachers' fidelity to the intended curriculum. These factors, however, may or may not negatively influence the intended scope of the curriculum and should be considered when making judgments about teacher fidelity. In conclusion, intervention researchers should consider the nature of the contextual factors and whether they negatively impact the intervention when designing and revising school-based interventions.

Keywords: fidelity, school-based physical education interventions, contextual factors, teacher interviews

Introduction

The American educational system has remained relatively stable for the last century as technology, personnel, and evidence-based practices have evolved in schools. Schools have primarily resisted changes in design, curricula, scheduling, and leadership structures (Seltz, 2008). However, the linear, factory-like ideals that historically facilitated school design are proving to be ineffective at meeting current students' needs. "Designed in another time, for the purposes of that time, the typical … school often shows a remarkable lack of flexibility" (Garmston & Wellman, 1995, p. 6). When the modern American educational system was designed in the early 1900s, approximately 51% of 5- to 19-year olds attended school (National Center for Education Statistics, 2012). Today's schools were never intended to meet the moral and economic demands of graduating all students. Furthermore, as the diversity of student populations increases, more poverty-stricken students will be entering America's schools. Tomorrow's students will demand curricula and teaching practices that offset the adverse living conditions that they have experienced since birth (Stevenson, 2010). Today's schools are complex from the classroom level to the district office level and are poorly represented by linear, factory-like frameworks (Davis & Sumara, 2005).

New school-based curricular interventions are fundamental in meeting the diverse needs of today's students and improving student outcomes. The No Child Left Behind Act of 2001 (NCLB; 2003) states that educators must use scientifically based teaching

methods and strategies proven to be effective. Proven means that programs have been evaluated through "scientifically based research … that involves the application of rigorous, systematic, and objective procedures to obtain valid knowledge relevant to education activities and programs" (U.S. Department of Education, 2002, p. 7). Interventions provide one method of measuring program effectiveness using rigorous, systematic, and objective procedures. Effective intervention designs use valid and reliable measures of achievement to compare intervention sites that use a particular program with sites that do not. If the intervention sites produce increased outcomes compared to the sites that do not implement the program, then the intervention/program may be labeled effective. However, the level of program effectiveness may be diminished or nullified if a program is implemented in a manner inconsistent with the innovator's plan. Therefore, measuring the extent to which a program implementer faithfully adheres (i.e., fidelity) to the innovator's program ideals during an intervention can contribute to program validation. According to Mowbray, Holter, Teague, and Bybee (2003), fidelity is the magnitude to which intervention delivery adheres to the program model originally developed. The measurement of fidelity during efficacy and effectiveness studies is gaining increased interest (National Research Council, 2004).

To maximize fidelity, researchers should identify contextual factors that hinder or facilitate interventions. According to Berman and McLaughlin (1976), three primary variables can impact school-based interventions and implementation fidelity: (a) federal policies, (b) institutional settings, and (c) project characteristics. Federal policies may mandate management theories and objectives, such as the requirement that

schools/students meet adequate yearly progress achievement outcomes. Due to adequate yearly progress standards, students may receive more instructional time in some disciplines compared to others, and as a result, impact an intervention's frequency and duration. Institutional setting factors, such as school and district characteristics, organizational climate, and staff motivation, can positively or negatively impact intervention effectiveness. Project characteristics, encompassing resource levels, the nature of the intervention, and implementation strategies, can increase or decrease intervention outcomes. At the institutional level, teachers are schools' main resource (Wayne & Youngs, 2003) and play a primary role in the faithful implementation of school-based interventions.

Public health and physical education pedagogy scholars have become increasingly concerned about the declining instructional time and overall quality of school-based physical education and physical activity programs. As a result, federal and foundational funding has become available to propose and test school-based physical activity interventions and disseminate effective alternatives to increase physical activity in youth. Research findings describing intervention characteristics and outcomes have been disseminated to inform scholars and practitioners of best practices. Yet, very few school-based physical activity intervention researchers have measured and reported fidelity levels (Sallis et al., 2012). Multiple organizations, such as the U.S. Department of Education, National Institutes of Health, and National Center on Response to Intervention, are requesting implementation fidelity research accompany large-scale intervention proposals. Recently, the National Institutes of Health (2011) established an

office of Dissemination and Implementation Research to support future implementation studies.

The purpose of this study was to investigate teachers' perceptions and justifications for changing critical intervention components and design principles. This research was guided by the following research question: what reasons do teachers offer for making changes to the intended curriculum as they implement it in their school contexts? When researchers understand teacher justifications for adapting intervention components, they can design or modify interventions to accommodate diverse contexts, minimizing factors that negatively impact implementation. The data gathered in this study provide information about the variables that enhanced or constrained the effectiveness of a large-scale physical education intervention and teacher rationales for intervention adaptations within their context.

Science of Healthful Living **Project**

The *Science of Healthful Living (SHL)* project was a 5-year National Institutes of Health funded project at the University of North Carolina at Greensboro. One goal of the *SHL* project was to develop a physically active, science-based middle school *SHL* curriculum. The *SHL* curriculum consists of two units: *Cardio Fitness Club* and *Healthy Lifestyles*. Each unit consists of twenty, 35-minute lessons using physical activity as a tool to teach health-related science and nutrition within the 5 Es (i.e., Engagement, Exploration, Explanation, Elaboration, and Evaluation) learning cycle strategy (Bybee et al., 1989). Students examine the effects of exercise on their bodies while participating in moderate to vigorous physical activities. Detailed rubrics, student science journals, and

validated multiple-choice tests provide opportunities for content valid student assessment. A detailed vocabulary list enhances opportunities for students to understand science terminology, concepts, and principles as they participate in enjoyable physical activities. Further information concerning the *SHL* project can be located at http://www.uncg.edu/hhs/science-of-healthful-living.

During Year 2, seven local education agencies from the Piedmont region of North Carolina participated in the *SHL* project. This included 25 middle schools (five schools were new to the project during Year 2), 16,000 diverse middle school students, and 70 middle school physical education teachers. Using free and reduced-price meals data and end-of-grade test scores (i.e., mathematics and reading), middle schools were stratified into matched pairs. Each school in each pair was then randomly assigned to teach either the *SHL* curriculum (experimental school) or traditional multi-activity physical education (control school). This study occurred during Year 2 as a part of the larger *SHL* project.

Implementation Fidelity

Because teachers implement instructional interventions in their classrooms, implementation fidelity hinges on teacher adoption. In the case of the *SHL* curriculum, I operationalized fidelity as the degree to which physical education teachers teach the lessons as structured while using effective teaching strategies. Threats to fidelity include factors that lead to the reduction of consistent delivery of the core elements of an intervention. In the case of the *SHL* curriculum, that meant the use of the 5 Es learning cycle strategy of science education (Bybee et al., 1989). Still, there are many other factors that can threaten implementation fidelity; therefore, researchers seek high levels

of standardization through the development of strategies such as program manuals, professional development training, and site visits by intervention supporters (Tyler & Blythe, 2008). Despite such efforts, researchers and teachers face a myriad of challenges to high implementation fidelity.

Implementation fidelity in school-based interventions can be measured using a number of strategies (O'Hare, 2005). Examples of strategies include case analyses, direct observations, process assessments, standardized questionnaires, and surveys. Measuring the implementation process gauges internal validity, that is, the extent to which the level of confidence researchers can assert that observable differences in outcomes – or lack thereof – were due to the intervention and not extraneous factors (Tucker & Blythe, 2008). Too few research studies in school settings, however, assess implementation fidelity (Tucker & Blythe, 2008). These findings underscore the need for more attention to implementation fidelity in educational research as a necessary component for establishing effective evidence-based practices in education (McLeod & Southam-Gerow, 2009; Mowbray et al., 2003; Smith, Daunic, & Taylor, 2007). Below, I describe several influences of implementation fidelity. Monitoring these factors can be effective for understanding participants' implementation methods, strengthening implementation fidelity protocols, and informing subsequent analyses and implementations.

Contextual Factors

In evaluating implementation fidelity, the implementation context often helps explain the degree to which implementation processes have an effect on intervention outcomes. In a meta-analysis of 81 studies, Durlak and DuPre (2008) identified 23

contextual factors that influenced implementation and situated those factors in three contextual categories: (á) community factors, (b) provider characteristics, and (c) intervention characteristics. Community factors affect an intervention by means of such dynamics as politics, funding, or policies. For example, the successful implementation of a multi-school intervention program may be conditional on school officials' ability to petition for funding of the intervention through political support. School officials include the superintendent, principals, teachers, and other school administrators who are in the position to facilitate or delay implementation.

The second contextual category includes teacher characteristics, such as perceptions of the intervention, beliefs about the need for the intervention, self-efficacy, teaching experience, and skill competence. For example, experienced teachers, already having years of teaching experience and feeling skilled with numerous teaching strategies, may be more reluctant to change their established teaching practices, thus affecting fidelity. However, compared to seasoned teachers, novice teachers may respond favorably to an intervention they assume may help strengthen their skills and add to their repertoire of teaching practices (Stein et al., 2008). Teachers with positive perceptions of an intervention may demonstrate confidence, consequently affecting their level of adherence and contribution to overall fidelity compared to teachers with a low opinion of the intervention (Durlak & DuPre, 2008). This suggests researchers' understanding of the range of teacher characteristics is an important dimension of intervention implementation.

The last contextual category, intervention characteristics, refers to the ease in which teachers can adapt the intervention to their needs and the extent to which the intervention is compatible with their roles, responsibilities, priorities, and the mission of the organization. An intervention that teachers consider "adaptable" suggests implementation fidelity is feasible when the components of the intervention complement their competencies. Overall, Durlak and DuPre (2008) suggest that fidelity may be attainable and sustained when intervention processes fit the organizational context, in that the components align with both teachers' characteristics and the established systems and structure in which teachers work.

Social Validity

Social validity is another intervention characteristic that contributes to fidelity. Social validity refers both to teachers' acceptance level of an intervention and their perceptions of intervention effectiveness in the classroom (Carter & Pesko, 2008). The relationship between social validity and implementation fidelity is an extension of Wolf's (1978) research on implementers' perceptions of an intervention and the intervention's social significance. Wolf (1978) explained that changes in behavior assessed by an objective measure (e.g., direct observation utilizing a reliable instrument) may contradict a teacher's report of program effectiveness. For example, a student in a treatment program may show slight improvement in academic performance, but feedback from the teacher may indicate significantly higher improvement. In this case, the difference between teacher feedback and the objective score underlines Wolf's contention that intervention effectiveness is contingent on the social validity of the teacher.

For teachers, one aspect of intervention feasibility is the extent to which the proposed intervention complements existing teaching practices (Carter & Pesko, 2008). Rather than competing for teachers' limited time, energy, and resources, or expecting them to exceed their roles and responsibilities, intervention practices that match classroom dynamics positively influence teachers' decisions to accept and adhere to a new intervention protocol (Carter & Pesko, 2008). Martens and McIntyre (2009) contend, however, that acceptability does not necessarily yield consistent and sustainable implementation of intervention protocols by teachers. To remedy declining fidelity among teachers, continuous support and reinforcement, monitoring, and feedback are suggested methods for ensuring implementation integrity over time (Martens & McIntyre, 2009).

Assessing teacher acceptability helps gauge the likely success of intervention implementation. According to Chafouleas, Briesch, Riley-Tillman, and McCoach (2009), practitioners' responses to an intervention as "appropriate, fair, and reasonable" can be assessed, especially when relevant, appropriate modifications to the intervention are undertaken to strengthen teacher acceptability and treatment effectiveness. Chafouleas and colleagues (2009) also noted, however, that attaining high levels of acceptability alone does not necessarily result in teachers' effective use of the intervention program or its components.

Research has confirmed that training and other forms of professional skill development are essential for ensuring fidelity, leading to the conclusion that there is a positive correlation between training and high levels of fidelity (Stead, Stradling,

Macneil, Mackintosh, & Minty, 2007). For example, Stead and colleagues (2007) concluded that time constraints in teachers' ability to deliver quality lessons, teachers' concerns about classroom disruptions during interactive assignments, and teachers' recognition of their limited content expertise were factors affecting their responses to the effectiveness of a new intervention (Sobeck, 2006). Further, program complexity appears to be associated with implementation fidelity, in that teachers may be less likely to implement complex interventions with high fidelity (Sobeck, 2006). In view of the challenges of maintaining implementation fidelity, in this study I investigated the reasons sixth-grade physical education teachers cite for changing or not following the structured lessons of the *SHL* curriculum.

Method

Participants

Purposeful sampling (Patton, 2002) was used to select six middle school physical education teachers participating for the first time in the larger *SHL* project. Teacher selection was based on their participation in the larger *SHL* project and on criteria (summarized below) that paralleled those identified by Hodge, Ammah, Casebolt, Lamaster, and O'Sullivan (2004) and Sato, Hodge, Murata, and Maeda (2007). First, all teachers were physical educators whose schools were located in adjacent school districts in the Piedmont region of North Carolina. Second, although these teachers were new to the *SHL* project, they had developed strong teaching records in their respective local education agency. Third, each had five or more years of teaching experience. According to Katz (1972), teachers are at the "maturity stage" once they meet these three conditions.

In the maturity stage, "teachers begin to ask questions of themselves and their teaching, focusing on insights, perspectives, and beliefs about teaching and children" (Stroot & Whipple, 2003, p. 313). Thus, participants in this research could reflect on their physical education teaching experiences. Because all six teachers participated in the larger *SHL* project, they were invited to 18 hours of professional development, received equipment to teach the lessons, and had a coach assigned to assist them in implementing the curriculum intervention in their schools. I served as the *SHL* coach for all six teachers in this study by providing intervention support and serving as a liaison between the teachers and the intervention developers. In this study, all six physical education teachers were teaching the sixth-grade *Cardio Fitness Club* unit of the *SHL* curriculum.

The selected teachers ranged in age from the upper twenties to the upper forties, had an average of 11 years ($SD = 6.6$) of teaching experience, and averaged 38.2 students ($SD = 15.3$) per sixth-grade physical education class. Table 5.1 presents the teachers' demographic and physical education class information.

Setting

Ms. East and Ms. Roberts taught middle school physical education to female students at Albany Middle School (pseudonym). Albany Middle School was an urban middle school with over 220 sixth-grade students (69% Black, 14% White, 7% Hispanic, 6% Asian, and 73% free and reduced-price meals [FARM]; Education First NC School Report Cards, 2011) during the 2012-2013 academic year. The school had one medium-sized gym and one small multipurpose room available for physical education instruction. Each sixth-grade physical education class was scheduled to meet for 38 minutes each

school day for the entire academic year. Ms. Roberts taught one sixth-grade physical education class and Ms. East taught two sixth-grade physical education classes. Ms. East's two sixth-grade physical education classes were combined to form one sixth-grade physical education class during data analysis.

Table 5.1

Teacher Demographics

Teacher	Gender	Age	Ethnicity	Years Teaching	Degree Level	Class Size
East	F	30-34	White	11	Masters	23
Roberts	F	25-29	Black	5	Masters	55
Barns	F	40-44	White	10	Bachelors	45
Parker	M	45-49	White	23	Bachelors	28
Adams	M	25-29	White	5	Bachelors	55
Nifong	F	35-39	White	12	Bachelors	23

Note. Teacher names are pseudonyms.

Ms. Barns and Mr. Parker taught coeducational middle school physical education at New River Middle School (pseudonym). New River Middle School was a city middle school with over 190 sixth-grade students (12% Black, 35% White, 48% Hispanic, and 77% FARM; Education First NC School Report Cards, 2011) during the 2012-2013 academic year. The school had one medium-sized gym available for physical education instruction. Each sixth-grade physical education class was scheduled to meet for 50 minutes each school day during the first semester. Ms. Barns and Mr. Parker both taught

sixth-grade physical education in the same gym during New River Middle School's third period. The teachers elected to team teach the *Cardio Fitness Club* lessons during third period. Ms. Barns was the lead teacher and Mr. Parker was the assisting teacher. Mr. Parker taught one sixth-grade physical education class during New River Middle School's fourth period and his class was the only class in the gymnasium during this period.

Mr. Adams and Ms. Nifong taught coeducational middle school physical education at St. Anthony Middle School (pseudonym). St. Anthony Middle School also was a city middle school with over 180 sixth-grade students (17% Black, 52% White, 25% Hispanic, and 60% FARM; Education First NC School Report Cards, 2011) during the 2012-2013 academic year. The school had one medium-sized gym available for physical education instruction. Each sixth-grade physical education class was scheduled to meet for 55 minutes each school day during the first semester. Mr. Adams and Ms. Nifong both taught sixth-grade physical education in the same gym during St. Anthony Middle School's first period. The teachers elected to team teach the *Cardio Fitness Club* lessons during first period. Mr. Adams was the lead teacher and Ms. Nifong was the assisting teacher. Ms. Nifong taught one sixth-grade physical education class during St. Anthony Middle School's second period and her class was the only class in the gymnasium during this period.

Data Collection

Once the teachers began teaching the sixth-grade *Cardio Fitness Club* unit, I began visiting each school. I began to collect detailed observation field notes on lesson

6. I did not collect observation field notes during lessons 1-5 to allow teachers time to adapt to the structure of the intervention lessons. I observed in a non-participatory capacity, comparing the lesson taught with the structured, scripted lesson that was provided in the intervention. I did not inform the teachers of when and which lessons I planned to observe. I documented the nature of the *Cardio Fitness Club* lessons for up to 5 weeks, observing six to seven lessons taught by each teacher. I paid special attention to the teacher's adherence and content delivery of the structured sixth-grade *Cardio Fitness Club* lessons.

After each lesson observation, I interviewed each teacher in their gymnasium asking unique questions based on the recently completed lesson observation. Each participating teacher was interviewed six to seven times during the study. The questions were based on observed deviations from the lessons as written and probed teachers on why they elected to make changes to the intended lessons.

Data Analysis

I analyzed the interview data through open, axial, and selective coding (Corbin & Strauss, 2008). During open coding I read, coded, and compared interview transcripts immediately following each school visit. This process "opened up the data" for interrogation, further inquiry, and conceptualization (Corbin & Strauss, 2008). The list of open codes referred to justifications teachers offered for lesson modifications. Once all transcripts were coded, I reviewed and reevaluated the coding for consistency and accuracy.

I used axial coding to organize the open coding categories into related clusters and generate connections among the open coding categories that explained the teachers' justifications. At this stage of analysis, interpretation of the data occurred as I identified relationships and linkages among open coding categories and attached meaning and significance to the analysis.

As themes emerged from the data, I used selective coding to connect these findings to the existing literature and further develop themes theoretically. During selective coding, categorical relationships defined during axial coding were connected and further refined to answer the research question. Central themes were related to each category through the use of explanatory relationship statements and then named and connected to the existing literature.

Credibility

Five primary techniques were utilized to maximize the credibility of the information gathered in this study. First, I spent 5 weeks interacting with the participants and observing their sixth-grade physical education classes. This time frame allowed me to gain the participants' trust and encouraged the participants to become comfortable with my presence in their classes. Second, I triangulated the data sources (i.e., lessons as written, detailed observation field notes, and interviews) to enhance the believability of the study (Patton, 2002). Third, I used member checks providing teachers with copies of their transcribed interviews and asking them to confirm and provide feedback regarding the accuracy of the transcriptions and any identified categories and themes (Creswell, 2011). Fourth, I wrote rich, thick descriptions of classroom observations to convey the

nature of the class context and the extent to which teachers followed the lesson plans provided by the intervention developers (Creswell, 2011). Fifth, I utilized a competent peer debriefer throughout data collection and analysis to maintain research ethics and maximize the trustworthiness of the data (Lincoln & Guba, 1985).

Results

The purpose of this study was to investigate teachers' perceptions and justifications for changing critical *SHL* intervention components and design principles. The findings of this study illuminated the effects of contextual factors on implementation fidelity. The concepts identified during the open and axial coding process were organized into seven categories that spoke to reasons why teachers changed lesson structure, tasks, or task order. One core category, school contextual constraints, emerged within this examination of the curriculum implementation process. The story line that follows is interlaced and interconnected with all six participants' voices.

School Contextual Constraints

Appropriate instructional time. Teachers' level of adaptability varied between implementing the intervention as intended and modifying the intervention to meet time constraints in the classroom. Teachers indicated two factors contributing to the variation of adaptability. In the first variation of adaptability, teachers discussed the length of their schools' physical education periods and the intended length of the *SHL* lessons. Albany, New River, and St. Anthony Middle Schools allotted 38, 50, and 55 minutes, respectively, for their physical education class periods.

For example, Ms. East at Albany Middle School expressed concern for the limited time available to implement the 35-minute lesson during her 38-minute physical education period:

> …38 minutes to teach a 35-minute lesson. Students need time to arrive to the gym, get their journals, and sit down. I need some time to introduce the lesson and teach the new activities. I had to skip the last two Es of the lesson today because we ran out of time.

Similarly, Ms. Roberts who also taught at Albany Middle School commented after a lesson by saying "I just don't have time to teach the whole lesson. I think I could do it if I had the time. As you saw today I barely started the third E [Explanation] and the students did not finish their journals." These two teachers (both from Albany Middle School) consistently referenced short physical education class periods as the main reason for modifying the intended lessons.

Both teachers attended the *SHL* professional development sessions where strategies were presented to manage physical education class periods less than or equal to 35 minutes. Ms. Roberts attended 12 out of the 18 hours offered and Ms. East attended all 18 hours. A few of the time-saving strategies presented by the *SHL* intervention team included reducing the length of the Engagement, disturbing journals by placing them on the bleachers so students could pick them up as they entered the gym, and not requiring students to change clothes for physical education. Ms. East and Ms. Roberts did not require their sixth-grade students to change clothes for physical education and consistently had journals laid out for students to pick up upon entering the gymnasium. They did not, however, elect to reduce the Engagement time.

The second variation of adaptability was related to the lessons having too much content to cover in the prescribed 35 minutes. Several participating teachers with 50- or 55-minute physical education periods tried to teach the *SHL* lessons in 35 minutes. These teachers attempted to motivate their students by offering "free play" or "student choice" opportunities in the class time remaining after the 35-minute *SHL* lesson. In some cases the teachers were successful, and in other cases content had to be omitted. For example, Mr. Parker noted

> This lesson [lesson 18] requires more than 35 minutes. I had to omit the explanation of ATP [adenosine triphosphate] and the physical activity homework to get through all the Es and journal work. I'm lucky that I have 50 minutes with this class so I can go over if I need to.

Along similar lines, a small group of participating teachers pointed out that the content was too challenging for sixth-grade students to learn in the allotted time. Ms. Nifong said "Energy systems are hard concepts for sixth graders. In the current structure, I don't have time to cover them well while still keeping the class moving." The teachers consistently questioned the amount of content to be covered in lessons 9, 16, and 18 from the sixth-grade *Cardio Fitness Club* unit. The topics of these three lessons were intensity of exercise, characteristics of anaerobic exercise, and energy systems, respectively.

Equipment needs. During my observations, I noticed participating teachers modifying or replacing the intended physical activities on a regular basis. One reason for these modifications was lack of equipment. All teachers participating in the larger *SHL* project received new, curriculum-specific equipment to teach the *SHL* curriculum. However, Mr. Adams explained during one interview "We only have 30 pedometers for

50 students. There was no way for us to give each student their own pedometer. We decided to let students work in pairs and share a pedometer." Furthermore, Ms. Barns shared "The department only has 20 jump ropes. It isn't enough for the whole class. I gave half the class a jump rope to use and told the other half to jump an imaginary jump rope. The lesson worked out to be the same." Jump ropes were not supplied by the *SHL* project, as it was assumed most middle physical education programs would have an ample supply.

Space needs. Another common reason for modifying the intended lessons was lack of space. All three schools participating in this study had at least one medium-sized gymnasium (i.e., large enough for a middle school basketball court and retractable bleachers). Albany Middle School also had a multipurpose room available for physical education instruction. Ms. Nifong said after lesson 18 that "Our gym isn't big enough to allow everyone to do the shuttle run at the same time. I decided to let the girls go first and then the boys." Additionally, Ms. East mentioned "We do not have the space to run four games of Steal the Diamonds at the same time. I had to cut it down to two games and rotated in new teams every 3 minutes." Ms. East, from Albany Middle School, was using the smaller multipurpose room the day she had to modify Steal the Diamonds.

Appropriate class sizes. Related to lack of equipment and lack of space is large class sizes. On a few occasions teachers justified their lesson modifications by pointing out their large class sizes. In this study, class sizes ranged from 23 students (Ms. East and Ms. Nifong) to 55 students (Mr. Adams). Ms. Roberts pointed out that "I had to add two stations to today's lesson [lesson 12]. The original lesson called for five stations but

that would've made the small groups too big and reduced physical activity time." Mr. Adams decided on a similar modification by adding a water station to reduce the number of students per group. During professional development, teachers were encouraged to add additional moderate to vigorous intensity stations to accommodate large classes in lessons utilizing station work.

Student skill level. Some teachers explained that low-skilled students were unable to perform skillfully at motor skill tasks as a reason to modify the lessons. A majority of the lessons in the *SHL* curriculum incorporated individualized fitness activities that could be differentiated for students' diverse fitness levels. Some of the lessons, however, utilized sports-based physical activities to help students learn content. Successful completion of sport-based activities requires a minimum skill level. Mr. Parker quickly noted during an interview that "... our sixth-grade students cannot do a [basketball] layup at all. I know the lesson [lesson 19] called for a layup station, but I decided to replace it with basketball wall passing." Ms. Roberts made a similar modification and explained after lesson 8 "... the students don't know how to throw a Frisbee. I decided to replace the Frisbee with a foam ball." Teachers were given the autonomy to make decisions like the above as long as the modified or new activities supported the original intent of the lessons.

Class management. One teacher voiced class management as a reason for modifying the intended lessons in the *SHL* curriculum. The teachers in this study utilized numerous class management techniques. For example, Ms. East expected her students to sit down along a specified wall anytime verbal instructions were to be communicated.

Mr. Parker used his whistle (i.e., one whistle blow) to notify students when it was time to stop moving and talking. During the observation of lesson 9 at New River Middle School, I noted that Mr. Parker replaced tennis racquet ball striking against a wall with badminton racquet shuttlecock hitting to a wall. After the lesson I asked Mr. Parker about the change and he explained "I changed that station to be safer. I didn't want tennis balls rolling all over the gym floor as other students were working at other stations. The shuttlecocks ain't going to go far after they bounce off the wall." On a different occasion I observed Mr. Parker control station rotation when the lesson provided students autonomy encouraging them to rotate to a new station when they finished their current station. Again I asked Mr. Parker about this modification and he pointed out "… students rotating on their own is too chaotic and didn't give equal opportunity to all the students at all the stations. I wanted to make sure most of the students had time to complete the work at each station before another group came." The *SHL* project staff encouraged teachers to manage their classes based on what they knew to be effective for their classes and school.

Teacher planning time. Teacher planning time was used as a reason for changing the lessons on a few occasions. On the average, Ms. Nifong and Ms. Roberts stated they spent less than 10 minutes planning to teach each lesson, Mr. Adams and Mr. Parker said they spent 10-20 minutes, and Ms. Barns and Ms. East said they spent greater than 20 minutes. During one observation (i.e., lesson 13) I observed Mr. Adams ask his students to perform two types of push-ups during the Engagement and two additional types of push-ups during the Elaboration. I noticed the two types of push-ups performed

during the Elaboration should have been performed during the Engagement. After the lesson, I asked Mr. Adams about the change and his response was "I messed that up pretty bad. I realized it [all four push-ups types were to be performed during the Engagement] when I got to the Elaboration. I arrived late to school this morning and didn't have time to look at the lesson." Teachers also were quick to note they had time to look at the lessons, but would put other responsibilities (i.e., coaching and athletic director) before reviewing and planning to teach the lessons. During the first 6-hour professional development, teachers were encouraged to read and study the lessons before implementing them. They were provided with recommendations on how to plan to teach the lessons and how long, on average, it would take to prepare to teach someone else's lessons.

Discussion

Weighing the Consequences of an Intervention Modification

School contextual factors have been shown to positively (e.g., Dariotis, Bumbarger, Duncan, & Greenberg, 2008; Elias, Zins, Graczyk, & Weissberg, 2003; Kam, Greenberg, & Walls, 2003; Wanless, Patton, Rimm-Kaufman, & Deutsch, 2012) and negatively (e.g., Kramer, Lauman, & Brunson, 2000; Wanless et al., 2012) promote implementation fidelity. In larger school-based interventions, such as randomized controlled trials, it is unlikely that all school contexts will lend themselves to implementation without adjustments or modifications. In these settings, developers expect modifications and monitor implementation to provide data for future intervention revisions. As discussed above there are different types of modifications that hold

different consequences for implementation fidelity. Some contextual modifications simply adjust the lessons for the unique characteristics of the school environment while others result in substantial changes to the intent of the intervention.

The type of modification should be considered when evaluating implementation fidelity. For example, little to no change occurred to the original intent of two lessons when Ms. Roberts from Albany Middle School decided to add two extra stations and replace a Frisbee with a foam ball. These types of changes were encouraged by the program staff to promote student success and did not negatively impact implementation fidelity. Conversely, the original intent of a lesson was negatively impacted when Mr. Parker from New River Middle School decided to reduce some of the content to be covered in a lesson even though he had ample time in his 50-minute period to fully cover the content. A change such as this had a negative impact on implementation fidelity.

Role of Teacher Attitudes and Beliefs

Additionally, interventions may require teachers to make choices that are inconsistent with their philosophical goals, attitudes, and beliefs concerning physical education. It is likely that decisions to implement tasks or policies in conflict can cause teacher distress and may lead to inconsistent or short-term adherence (Beets et al., 2008; Evans, 1996; Ringwalt et al., 2003; Rohrbach, Graham, & Hansen, 1993). Teachers' may demonstrate less resistance to a new intervention when (a) there is no external requirement of implementation (Smylie, 1988), (b) teachers have autonomy in program adoption decisions (Parcel et al., 1991), and (c) appealing professional development is made available for teachers' at their convenience (Kent, 2004).

As described in the results section, class management and teacher planning time were two teacher justifications for making changes to the intended curriculum. These justifications are related to and stem from teachers' beliefs and attitudes. For example, Mr. Parker from New River Middle School had concerns about students' safety and equal access to physical activities and made changes to the intended lessons based on his beliefs about effective class management. Likewise, Mr. Adams from St. Anthony Middle School elected not to plan and prepare for an appropriate amount of time before some lessons. Mr. Adams based this decision on his priorities, which ranked other school-based responsibilities in front of preparing to teach the intervention lessons. Mr. Adams' decision-making process was at least partially mediated by his beliefs about physical education and attitudes toward the *SHL* curriculum.

Both New River Middle School and St. Anthony Middle School were in the same local education agency where a district administrator decided that the physical education teachers at New River and St. Anthony Middle Schools would implement the *SHL* curriculum. It is unclear what impact this top-down decision had on Mr. Parker's and Mr. Adam's acceptance of the intervention, but it is clear that neither teacher had a role in intervention adoption. Furthermore, both teachers were invited to and attended 3 days (18 hours) of project-sponsored professional development. The *SHL* project team secured a centralized meeting location, organized and led the sessions, provided breakfast and lunch, paid for teacher substitutes, and awarded continuing education units that could be used for renewing professional teaching licenses as required by the state of North Carolina. Some of the professional development included class management techniques

and teacher preparation suggestions. Given the nature of this study, it is not known what impact the professional development had on these two teachers' attitudes towards and beliefs about the *SHL* curriculum. The project staff, however, attempted to create professional development opportunities for all participating teachers that were both convenient and appealing in hopes of promoting teacher buy-in.

Measuring teacher fidelity levels are helpful to ensure that outcomes are robust and valid in intervention research (Tucker & Blythe, 2008). As previously indicated, there are several methods to evaluate fidelity including randomized observational checks, videotaped intervention sessions, audio taped sessions, process assessment, standardized evaluation forms, and post-session assessment. Such methods help evaluators and scholars organize and categorize barriers to teacher fidelity, identify the effectiveness of program components, incorporate practitioners' suggestions, and distinguish variations found in program outcomes.

Conclusions

In the current study, I used qualitative methods to examine teachers' reasons for modifying structured lessons from the *SHL* curriculum. The current findings suggest teachers changed the lessons due to instructional time, space, equipment, class size, student skill level, class management concerns, and teacher planning time. The current findings show potential for informing scholars of the merit and efficacy of the intervention on student outcomes. Thus, research on implementation fidelity has the potential to advance the quality of evidence-based program development and research. Further, growing demands to demonstrate the efficacy and scalability of school reform

programs should drive the need for fidelity evaluations to identify variations of program effectiveness across different school-based populations. For that reason, research on implementation fidelity adds value to existing services, especially when it helps practitioners' enhance their knowledge and skills, thus better standardizing intervention delivery.

CHAPTER VI

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

Summary

The American idea of schools has remained fairly stable since the state of Massachusetts first offered its citizens a free public education over 150 years ago. Today's students still attend schools with age-based grade levels, rigid credit requirements, bell schedules, and historically facilitated physical characteristics. These factory-model schools, however, do not appear to be educating its students to be productive members of an ever-changing society. The schools of Horace Mann's time were neither designed to educate all students, nor prepare them for a college education. The factory-like schools frequently seen across the United States are poor representations of the complex mechanisms at work in today's classrooms and school and district administrative offices.

As researchers continue to design, implement, and evaluate school-based interventions to better serve students, teachers, and schools, complexity theory may serve as a framework for better understanding today's educational system (Fullan, 2008), because it addresses persevering patterns, evidenced through history and impervious to change. Complexity theory can function as a theoretical framework for explaining and understanding school-based interventions and recurring themes appearing in schools. Complexity theory describes multiple levels of interactions that result in opportunities

and constraints to the change process within organizations, such as schools (Richardson & Cilliers, 2001).

New school-based curricular interventions are fundamental in meeting the diverse needs of today's students and improving student outcomes. Interventions provide one method of measuring program effectiveness using rigorous, systematic, and objective procedures. The level of program effectiveness, however, may be diminished or nullified if a program is implemented in a manner inconsistent with the innovator's plan. Therefore, measuring the extent to which a program implementer faithfully adheres (i.e., fidelity) to the innovator's program ideals during an intervention can contribute to program validation. According to Mowbray, Holter, Teague, and Bybee (2003), fidelity is the magnitude to which intervention delivery adheres to the intervention model initially developed.

Interventions are typically designed for implementation in limited contexts with the purpose of increasing programmatic efficacy and internal validity. Once the intervention is disseminated, however, the implementation conditions may be different, or the manner of implementation may vary from the original design. Fidelity measures offer a connection between implementation and outcomes to better understand program effectiveness (Sanchez et al., 2007). Fidelity of implementation and program effectiveness is directly related. Faithfully implemented interventions are likely to produce superior student outcomes compared to those that are not (Burns, Peters, & Noell, 2008; Durlak & DuPre, 2008). In the absence of fidelity measures, intervention designers and implementers may not know if the intervention is effective and if so, in

what environments. Thus, measuring fidelity, as part of research-based interventions, promotes validity and can explain variance between outcomes.

Public health and pedagogy scholars have become increasingly concerned about the declining instructional time and overall quality of school-based physical education and physical activity programs. As a result, federal and foundational funding has become available to propose and test school-based physical activity interventions and disseminate effective alternatives to increase physical activity in youth. Research findings describing intervention characteristics and outcomes have been disseminated to inform scholars and practitioners of best practices. Yet, very few school-based physical activity intervention researchers have measured and reported fidelity levels (Sallis et al., 2012).

Measuring fidelity and examining its impact on student outcomes is a promising concept in educational research (Mowbray et al., 2003; O'Donnell, 2008). The purpose of this study was to examine sixth-grade physical education teachers' fidelity levels when implementing the *Cardio Fitness Club* unit from the *Science of Healthful Living* curriculum and to investigate the relationship between teacher fidelity levels and student outcomes (i.e., knowledge growth and physical activity intensity levels). I also investigated the extent to which teachers adapted or ignored critical intervention components and design principles and teachers' perceptions and justifications for these changes as they influence fidelity level.

Assessing implementation fidelity increases researchers' and practitioners' understanding of the fit between the intervention and the context. However, the complexity of the implementation context can impact the efficiency and generalizability

of the innovation to other situations. Researchers have rarely used fidelity variables to explain intervention outcomes. Appropriate and accurate fidelity data can inform research and practice to better explain intervention outcomes. Further, when researchers understand teacher justifications for adapting intervention components, they can design or modify interventions to accommodate diverse contexts, minimizing factors that negatively impact implementation.

During a time of teacher and student accountability, high stakes testing, and No Child Left Behind mandates, schools and students must have effective curricula to increase achievement outcomes. Furthermore, given the calls by public health organizations to increase physical activity and decrease overweight and obesity levels in children and adolescents, school-based physical education programs need evidence-based interventions proven to increase educational achievement and /or physical activity outcomes. Measuring implementation fidelity provides one method of assessing moderators of curricular effectiveness. The data gathered in this study provided information about the variables that enhanced or constrained the effectiveness of a large-scale physical education intervention, the influence of implementation fidelity on student outcomes, and teacher rationales for intervention adaptations within their context.

I used a mixed methods design to collect and analyze both quantitative and qualitative data examining implementation fidelity in the *Science of Healthful Living* project (Creswell, 2013; Creswell & Plano Clark, 2011; Tashakkori & Teddlie, 2010). Using multiple data collection strategies provided a more complete analysis than any single method alone could accomplish when addressing the problem (Creswell, 2013;

Creswell & Plano Clark, 2011; Tashakkori & Teddlie, 1998; 2010). A mixed methods design was appropriate because fidelity of implementation measurement and analysis required both quantitative and qualitative data collection and analysis during various study stages. Mixed methods approaches offered complementary insights into understanding the faithful implementation of the *Science of Healthful Living* curriculum (RQ1), the impact fidelity of implementation had on student outcomes (RQ2), and the reasons teachers cited for making modifications to the intended curriculum (RQ3).

Participants were middle school physical education teachers and their sixth-grade physical education students from Albany, New River, St. Anthony, and Titan Middle Schools (pseudonyms). These middle schools were located in the Piedmont region of NC and served students in Grades 6-8; ages 11-14. Each school was selected because the teachers were voluntarily participating in the *Science of Healthful Living* project effectiveness study for the first time during the 2012-2013 school year. The four schools selected for this research had been randomly assigned to the intervention group. Physical education teachers taught the *Science of Healthful Living* curriculum consisting of the 20 structured lessons from the sixth-grade *Cardio Fitness Club* unit.

Depending on school academic calendars and school computer lab availability, participating teachers administered the sixth-grade, 10-item standardized knowledge pretest to their sixth-grade physical education students prior to teaching *Cardio Fitness Club* lesson 1. Once knowledge pretesting was complete at each participating school, participating teachers began teaching the sixth-grade *Cardio Fitness Club* unit (lessons 1-20) to their sixth-grade students. I began visiting each participating school beginning

with lesson six to collect detailed field notes on lessons 6-20. During lessons 6-20 of the sixth-grade *Cardio Fitness Club* unit, I also collected student physical activity intensity level data each time I observed a lesson and interviewed the observed teacher asking unique questions based on the recently completed lesson observation. After completing the 20 lessons, participating teachers re-administered the 10-item test used for pretesting to their sixth-grade students.

Open and axial coding was applied to the lesson observations to develop a 42-item dichotomous rubric to quantify each teacher's fidelity level. Teacher interview transcripts were coded using open, axial, and selective coding to produce categories and themes. Knowledge test results were converted to an average growth score per teacher and student physical activity intensity levels were converted to an average vector magnitude count per minute per teacher. Multiple regression was used to assess the contribution of implementation fidelity to the prediction of class knowledge change and class physical activity intensity levels over the course of the 20-lesson *Cardio Fitness Club* unit.

The findings of this study illuminated the effects of contextual factors on implementation fidelity. The concepts identified during the open, axial, and selective coding of the teacher interviews were organized into seven categories that spoke to reasons why teachers changed scripted lessons. One core category, school contextual constraints, emerged within this examination of the curriculum implementation process. Overall fidelity score accounted for 79% of the variance in knowledge change. When the 4 Es were entered as predictors, Engagement, Exploration, Explanation, and Evaluation

contributed to the overall model fit for knowledge change. The beta coefficients for Engagement and Explanation were statistically significant. Multiple regression targeting class-level physical activity intensity levels did not produce any significant findings using overall fidelity score or the 4 Es' scores.

In larger school-based interventions, such as randomized controlled trials, it is unlikely that all school contexts will lend themselves to implementation without adjustments or modifications. In these settings, developers expect modifications and monitor implementation to provide data for future intervention revisions. Some contextual modifications simply adjust the lessons for the unique characteristics of the school environment while others result in substantial changes to the intent of the intervention. The type of modification should be considered when evaluating implementation fidelity.

Multiple regression showed that overall fidelity of implementation accounted for 79% of the variance in fitness-based knowledge growth in sixth-grade students participating in the curricular intervention. Overall fidelity of implementation and specific Es (i.e., Engagement and Explanation) made a statistically significant contribution to the prediction of student knowledge gain. These findings reinforce earlier research demonstrating that fidelity of implementation has statistically significant implications on student outcomes (Allinder et al., 2000; Butler-Songer & Gotwals, 2005; Hall & Loucks, 1977; Penuel & Means, 2004; Ysseldyke et al., 2003).

Conclusions

This dissertation research was intended to answer three research questions:

1. To what extent do teachers implement lessons from a sixth-grade physical education curricular intervention with fidelity (adherence – coverage and delivery quality)?
2. What is the relationship between implementation fidelity and student outcomes (i.e., knowledge gain and physical activity intensity levels)?
3. What reasons do teachers offer for making changes to the intended curriculum as they implement it in their school contexts (moderating variable of context)?

Major findings that answered these research questions include the following:

1. Overall fidelity scores ranged from 42% to 83% across the six teachers implementing the sixth-grade *Cardio Fitness Club* unit from the *Science of Healthful Living* curriculum. Additionally, four of the six overall fidelity scores were 75% or higher (see Table 4.2).
2. Overall fidelity score accounted for 79% [adj $R^2 = .74$, $F(1, 4) = 15.4$, $p < .05$] of the variance in knowledge change. The overall fidelity beta coefficient was statistically significant (see Table 4.4). When the 4 Es were entered as predictors, Engagement, Exploration, Explanation, and Evaluation contributed to the overall model fit in the regression analysis for knowledge change [$R^2 = .99$, adj $R^2 = .99$, $F(4, 1) = 416.99$, $p < .05$]. The beta coefficients for Engagement and Explanation were statistically significant (see Table 4.4). Multiple regression targeting class-

level physical activity intensity levels did not produce any significant findings using overall fidelity score or the 5 Es' scores.

3. Seven categories related to teacher intentions to modify the curriculum emerged from the analysis: (a) appropriate instructional time, (b) equipment needs, (c) space needs, (d) appropriate class sizes, (e) student skill level, (f) class management, and (g) teacher planning time.

Collectively the findings indicated that (a) overall fidelity scores across teachers implementing the same curriculum and receiving the same professional development and support from the implementation team can vary across schools and teachers. Multiple preexisting school contextual factors impacted the variation in overall fidelity scores. (b) Students learned more as teachers taught lessons with more fidelity. The teachers with higher overall fidelity scores taught physical education classes demonstrating more knowledge gain. (c) Teachers stated appropriate instructional time, equipment needs, space needs, appropriate class sizes, student skill level, class management, and teacher planning time initiated their changes to the intended curriculum. These rationales may have constrained overall fidelity scores, which in turn may have limited student knowledge gain.

Recommendations

The research findings from this dissertation have significant theoretical implications for physical education research and lead me to suggest several recommendations for future research. First, teachers implementing the *Science of Healthful Living* curriculum (a research-based physical education intervention) with

lower fidelity of implementation experienced smaller increases in knowledge gain compared to teachers implementing the intervention with higher fidelity of implementation. Failing to consider fidelity of implementation is a major threat to internal validity. The intervention's role in enhancing or constraining outcomes cannot be fully determined without measuring fidelity of implementation. Given that few studies to date have assessed the role of fidelity of implementation on physical education outcomes, future fidelity research in physical education is justified. Monitoring fidelity of implementation may modify outcomes of existing and future program effectiveness, efficacy, and evaluation studies of physical education curricula.

Second, I recommend that future fidelity evaluators collect systemic data (i.e., observations and teacher interviews) early in the design process to describe the current pre-intervention instructional process and procedures being used in the physical education classes that will be the intervention targets. Physical education teachers in this study noted several preexisting contextual factors impacting their ability to teach the lessons as intended. Lack of instructional time, space, and equipment, large class sizes, and student skill level were noted as some of the reasons for changing the intended lessons. Preexisting contextual factors need to be considered during the design process, as these factors may not be open for change during implementation. When problems in adherence or fidelity are found, immediately collect additional data from teachers to better understand the sources of their challenges; then, when possible, make program changes to the intervention to address the issues.

Third, in any teacher implemented intervention study, understanding the nature of traditional curricula is essential. Assessing the usual curriculum prior to an instructional intervention implementation and evaluation is uncommon. Measuring routine teacher behavior as an element of the general intervention work is costly. Still, such procedures are important because understanding similarities and differences between pre-intervention teaching processes and new intervention processes may help educational program developers and scholars better understand the degree to which a new instructional intervention a) overlaps with current curricula, b) exceeds or deviates from practitioners' roles and responsibilities, or c) undermines targeted teacher knowledge, skills, or supports.

REFERENCES

Actigraph. (2012, June). ActiLife 6 Data Analysis Software (Version 6.2) [Computer Software]. Pensacola, FL: Actigraph. Available from http://www.actigraphcorp.com/

Alexandrov, A. A., Maslennikova, G. Y., Kulikov, S. M., Propirnij, G. A., & Perova, N. V. (1992). Primary prevention of cardiovascular disease: 3-year intervention results in boys of 12 years of age. *Preventive Medicine, 21,* 53-62. Retrieved from http://www.journals.elsevier.com/preventive-medicine/

Allinder, R. M., Bolling, R. M., Oats, R. G., & Gagnon, W. A. (2000). Effects of teacher self-monitoring on implementation of curriculum-based measurement and mathematics computation achievement of students with disabilities. *Remedial and Special Education, 21*, 219-226. doi:10.1177/074193250002100403

Axelrod, R., & Cohen, M. D. (2000). *Harnessing complexity.* New York, NY: Basic Books.

Bachrach, L. L. (1988). The chronic patient: On exporting and importing model programs. *Hospital and Community Psychiatry, 39*, 1257-1258. Retrieved from http://ps.psychiatryonline.org/issues.aspx?journalid=18

Barbeau, P., Johnson, M. H., Howe, C. A., Allison, J., Davis, C. L., Gutin, B., & Lemmon, C. R. (2007). Ten months of exercise improves general visceral

adiposity, bone, and fitness in Black girls. *Obesity, 15,* 2077-2085. doi:10.1038/oby.2007.247

Basch, C., Sliepcevich, E., Gold, R., Duncan, D., & Kolbe, L. (1985). Avoiding type III errors in health education program evaluations: A case study. *Health Education Quarterly, 54*, 362-395. doi:10.1177/109019818501200311

Beets, M. W., Flay, B. R., Vuchinich, S., Acock, A. C., Li, K., & Allred, C. (2008). School climate and teachers' beliefs and attitudes associated with implementation of the positive action program: A diffusion of innovations model. *Prevention Science, 9*, 264-275. doi:10.1007/s11121-008-0100-2

Berman, P., & McLaughlin, M. (1976). Implementation of educational innovation. *The Educational Forum, 40*, 345-370. doi:10.1080/00131727609336469

Blakely, C., Mayer, J., Gottschalk, R., Schmitt, N., Davidson, W., Roitman, D., Emshoff, J. (1987). The fidelity-adaptation debate: Implications for the implementation of public sector social programs. *American Journal of Community Psychology, 15*, 253-268. Retrieved from http://www.springer.com/psychology/community+psychology/journal/10464

Borman, G., Slavin, R. E., Cheung, A., Chamberlin, A., Madden, N. A., & Chambers, B. (2007). Final reading outcomes of the national randomized field trail of Success for All. *American Educational Research Journal, 44*, 701-731. doi:10.3102/0002831207306743

Borrelli, B., Sepinwall, D., Bellg, A. J., Breger, R., DeFrancesco, C., Sharp,...Orwig, D. (2005). A new tool to assess treatment fidelity and evaluation of treatment fidelity

across 10 years of health behavior research. *Journal of Consulting and Clinical Psychology, 73*, 852-860. doi:10.1037/0022-006X.73.5.852

Brunette, M., Asher, D., Whitley, R., Lutz, W., Wieder, B., Jones, A., & McHugo, G. (2008). Implementation of integrated dual disorders treatment: A qualitative analysis of facilitators and barriers. *Psychiatric Services, 59*, 989-995. Retrieved from http://ps.psychiatryonline.org/journal.aspx?journalid=18

Burns, M. K., Peters, R., & Noell, G. H. (2008). Using performance feedback to enhance implementation fidelity of the problem-solving team process. *Journal of School Psychology, 46*, 537-550. doi:10.1016/j.jsp.2008.04.001

Butler-Songer, N., & Gotwals, A. W. (2005, April). *Fidelity of implementation in three sequential curricular units.* In S. Lynch (Chair) & C. L. O'Donnell, *Fidelity of implementation in implementation and scale-up research designs: Applications from four studies of innovative science curriculum materials and diverse populations.* Symposium conducted at the Annual Meeting of the American Educational Research Association, Montreal, Canada. Retrieved from http://www.gwu.edu/~scale-up/fidelity.html

Bybee, R. W., Buchwald, C. E., Crissman, S., Heil, D. R., Kuerbis, P. J., Matsumoto, C., & McInerney, J. D. (1989). *Science and technology education for the elementary years: Frameworks for curriculum and instruction.* Washington, DC: The National Center for Improving Science Education. Retrieved from ERIC database. (ED314237)

Caballero, B., Clay, T., Davis, S. M., Ethelbah, B., Rock, B. H., Lohman, T., . . . Stevens, J. (2003). Pathways: A school-based, randomized controlled trial for the prevention of obesity in American Indian school children. *American Journal of Clinical Nutrition, 78,* 1030-1038. Retrieved from http://ajcn.nutrition.org/

Capra, F. (1996). *The web of life. A new synthesis of mind and matter.* London, UK: HarperCollins.

Capra, F. (2002). *The hidden connections*. London, UK: HarperCollins.

Carroll, C., Patterson, P., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science, 2*(40), 1-9. doi:10.1186/1748-5908-2-40

Carter, E., & Pesko, M. (2008). Social validity of peer interaction intervention strategies in high school classrooms: Effectiveness, feasibility, and actual use. *Exceptionality, 16*(3), 156-173. doi:10.1080/09362830802198427

Century, J., Rudnick, M., & Freeman, C. (2010). A framework for measuring fidelity of implementation: A foundation for shared language and accumulation of knowledge. *American Journal of Evaluation, 31*, 199-218. doi:10.1177/1098214010366173

Chafouleas, S., Briesch, A., Riley-Tillman, T., & McCoach, B. (2009). Moving beyond assessment of treatment acceptability: An examination of the factor structure of the Usage Rating Profile-Intervention (URP-I). *School Psychology Quarterly, 24*(1), 36-47. doi:10.1037/a0015146

Chen, A., Martin, R., Sun, H., & Ennis, C. D. (2007). Is in-class physical activity at risk in constructivist physical education? *Research Quarterly for Exercise and Sport, 78*, 500-509. doi:10.1080/02701367.2007.10599449

Chen, H. (1990). *Theory-driven evaluations.* Thousand Oaks, CA: Sage.

Chow, B. C., McKenzie, T. L., & Louie, L. (2009). Physical activity and environmental influences during secondary school physical education. *Journal of Teaching in Physical Education, 28,* 21-37. Retrieved from http://journals.humankinetics.com/JTPE

Clancy, T. R. (2004). Navigating in a complex nursing world. *Journal of Nursing Administration, 34*, 274-282. Retrieved from http://journals.lww.com/jonajournal/pages/default.aspx

Colchico, K., Zybert, P., & Basch, C. E. (2000). Effects of after-school physical activity on fitness, fatness, and cognitive self-perceptions: A pilot study among urban, minority adolescent girls. *American Journal of Public Health, 90,* 977-978. doi: 10.2105/AJPH.90.6.977

Cook, B. G., Landrum, T. J., Tankersley, M., & Kauffman, J. M. (2003). Bringing research to bear on practice: Effecting evidence-based instruction for students with emotional or behavioral disorders. *Education and Treatment of Children, 26*, 345-361. Retrieved from http://www.educationandtreatmentofchildren.net/index.html

Corbin, J., & Strauss, A. (2008). *Basics of qualitative research* (3rd ed.). Thousand Oaks, CA: Sage.

Coveney, P., & Highfield, R. (1995). *Frontiers of complexity: The search for order in a chaotic world.* New York, NY: Fawcett Columbine.

Creswell, J. W. (2011). *Educational research: Planning, conducting, and evaluating quantitative and qualitative approaches to research* (4th ed.). Upper Saddle River, NJ: Pearson Education.

Creswell, J. W. (2013). *Qualitative inquiry and research design: Choosing among five approaches* (3rd ed.). Thousand Oaks, CA: Sage.

Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research* (2nd ed.). Thousand Oaks, CA: Sage.

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation efforts out of control? *Clinical Psychology Review, 18*, 23-45. Retrieved from http://www.sciencedirect.com/science/journal/02727358

Dariotis, J. K., Bumbarger, B. K., Duncan, L. G., & Greenberg, M. T. (2008). How do implementation efforts relate to program adherence? Examining the role of organizational, implementer, and program factors. *Journal of Community Psychology, 36*, 744-760. doi:10.1002/jcop.20255

Darling-Hammond, L., Berry, B., & Thomerson, A. (2001). Does teacher certification matter? Evaluating the evidence. *Educational Evaluation and Policy Analysis, 23* (1), 57-77. Retrieved from http://www.jstor.org/stable/3594159

Datnow, A., & Castellano, M. (2000). Teachers' responses to Success for All: How beliefs, experiences, and adaptations shape implementation. *American Educational Research Journal, 37*, 775-799. doi:10.3102/00028312037003775

Davis, B., & Sumara, D. J. (2005). Challenging images of knowing: Complexity science and educational research. *International Journal of Qualitative Studies in Education, 18*, 305-321. doi:10.1080/09518390500082293

Dobson, L., & Cook, T. (1980). Avoiding Type III error in program evaluation: Results from a field experiment. *Evaluation and Program Planning, 3*, 269-276. Retrieved from http://www.journals.elsevier.com/evaluation-and-program-planning/

Dochy, F., Segers, M., & Buehl, M. M. (1999). The relation between assessment practices and outcomes of studies: The case of research on prior knowledge. *Review of Educational Research, 69*, 145-186. doi:10.3102/00346543069002145

Domitrovich, C., & Greenberg, M. (2000). The study of implementation: Current findings from effective programs that prevent mental disorder in school-aged children. *Journal of Educational and Psychological Consultation, 11*, 193-221. doi:10.1207/S1532768XJEPC1102_04

Donnelly, J. E., Jacobsen, D. J., Whatley, J.E., Hill, J. O., Swift, L. L., Cherrington, A., . . . Reed, G. (1996). Nutrition and physical activity program to attenuate obesity and promote physical and metabolic fitness in elementary school children. *Obesity Research, 4,* 229-243.

Dorland, W. (1994). *Dorland's illustrated medical dictionary* (28th ed.). Philadelphia, PA: W. B. Saunders.

Drake, R., Goldman, H., Leff, H., Lehman, A., Dixon, L., Mueser, K., & Torrey, W. (2001). Implementing evidence-based practices in routine mental health service settings. *Psychiatric Services, 52*, 179-182. Retrieved from http://ps.psychiatryonline.org

Dumas, J. E., Lynch, A. M., Laughlin, J. E., Smith, E. P., & Prinz, R. J. (2001). Promoting intervention fidelity: Conceptual issues, methods, and preliminary results from the EARLY ALLIANCE prevention trial. *American Journal of Preventive Medicine, 20*(1; Suppl. 1), 38-47. doi:10.1016/S0749-3797(00)00272-5

Durlak, J. A. (1998). Why program implementation is important. *Journal of Prevention and Intervention in the Community, 17*, 5-18. doi:10.1300/J005v17n02_02

Durlak, J., & DuPre, E. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology, 41*, 327-350. doi:10.1007/s10464-008-9165-0

Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. B. (2003). A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings. *Health Education Research, 18*, 237-256. doi:10.1093/her/18.2.237

Dusenbury, L., Brannigan, R., Hansen, W. B., Walsh, J., & Falco, M. (2005). Quality of implementation: Developing measures crucial to understanding the diffusion of

preventive interventions. *Health Education Research, 20*, 308-313. doi:10.1093/her/cyg118

Education First NC School Report Cards. (2011, October). *NC school report cards – 2010-2011 school year.* Retrieved from www.ncreportcards.org

Elias, M. J., Zins, J. E., Graczyk, P. A., & Weissberg, R. P. (2003). Implementation, sustainability, and scaling up of social-emotional and academic innovations in public schools. *School Psychology Review, 32*, 303-319. Retrieved from http://www.nasponline.org/publications/spr/index-list.aspx

Elliott, D., & Mihalic, S. (2004). Issues in disseminating and replicating effective prevention programs. *Prevention Science, 5*, 47-53. Retrieved from http://www.springer.com/public+health/journal/11121

Elliot, E., & Kiel, L. (1997). Nonlinear dynamics, complexity, and public policy. In R. Eve, S. Horsfall, & M. Lee (Eds.), *Chaos, complexity, and sociology: Myths, models, and theories* (pp. 64-78). Thousand Oaks, CA: Sage.

Ennis, C. D. (1992). Reconceptualizing learning as a dynamical system. *Journal of Curriculum and Supervision, 7*, 115-130. Retrieved from http://www.ascd.org/publications/jcs/archived-issues.aspx

Ennis, C. D. (1995) Teachers' responses to noncompliant students: The realities and consequences of a negotiated curriculum. *Teaching and Teacher Education, 11*, 445-460. Retrieved from http://www.journals.elsevier.com/teaching-and-teacher-education/

Ennis, C. D. (1999). Creating a culturally relevant curriculum for disengaged girls. *Sport, Education, and Society, 4*, 31-49. doi:10.1080/1357332990040103

Evans, R. (1996). The human side of school change: Reform, resistance, and the real-life problems of innovation. San Francisco, CA: Jossey-Bass.

Everhart, K., & Wandersman, A. (2000). Applying comprehensive quality programming and empowerment evaluation to reduce implementation barriers. *Journal of Educational and Psychological Consultation, 11*, 177-191. doi:10.1207/S1532768XJEPC1102_03

Field, A. (2009). *Discovering statistics using IBM SPSS Statistics* (3rd ed.). Thousand Oaks, CA: Sage.

Freedson, P., Pober, D., & Janz, K. F. (2005). Calibration of accelerometer output for children. *Medicine & Science in Sports & Exercise, 37*(Suppl. 11), S523-S530. doi:10.1249/01.mss.0000185658.28284.ba

Fullan, M. (1983). Evaluating program implementation: What can be learned from follow through. *Curriculum Inquiry, 13*, 215-227. Retrieved from http://www.jstor.org/stable/1179640

Fullan, M. (2003). *Change forces with a vengeance.* London, UK: Routledge Falmer.

Fullan, M. (2008). *What's worth fighting for in the principalship?* (2nd ed.). New York, NY: Teachers College Press.

Fullan, M., & Pomfret, A. (1977). Research on curriculum and instruction implementation. *Review of Educational Research, 47*, 335-397. doi:10.3102/00346543047002335

Garmston, R., & Wellman, B. (1995, April). Adaptive schools in a quantum universe. *Educational Leadership, 52*(7), 6-12. Retrieved from http://www.ascd.org/publications/educational-leadership/apr95/vol52/num07/Adaptive-Schools-in-a-Quantum-Universe.aspx

Gersten, R., & Brengelman, S. U. (1996). The quest to translate research into classroom practice: The emerging knowledge base. *Remedial and Special Education, 17*, 67-74. doi:10.1177/074193259601700202

Glaser, B. (1978). *Theoretical sensitivity: Advances in the methodology of ground theory.* Mill Valley, CA: Sociology Press.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research.* New York, NY: Aldine.

Gleick, J. (1987). *Chaos: Making a new science.* London, UK: Sphere Books.

Goldhaber, D., & Brewer, D. (1996). Evaluating the effect of teacher degree level of educational performance. In W. Fowler (Ed.), *Developments in school finance* (pp.197-210). Washington, DC: National Center for Education Statistics.

Goldstein, J. (2008). Resource guide and glossary for non-linear/complex systems terms. In C. Linberg, S. Nash, & C. Lindberg (Eds.), *On the edge: Nursing in the age of complexity* (pp. 263-294). Bordentown, NJ: Plexus Press.

Gortmaker, S. L., Cheung, L. W., Peterson, K. E., Chomitz, G., Cradle, J. H., Dart, H, . . . Laird, N. (1999). Impact of a school-based interdisciplinary intervention on diet and physical activity among urban primary school children: Eat well and keep

moving. *Archives of Pediatrics & Adolescent Medicine, 153,* 975-983. Retrieved from http://archpedi.jamanetwork.com/journal.aspx

Greenberg, M. T., Domitrovich, C. E., Graczyk, P. A., & Zins, J. E. (2005). *The study of implementation in school-based preventative interventions: Theory, research, and practice. (Volume 3).* Rockville, MD: Center for Mental Health Services, Substance Abuse and Mental Health Services Administration.

Greenhalgh, T., Robert, G., Macfarlane, F., Bate, P., & Kyriakidou, O. (2004). Diffusion of innovations in service organizations: Systematic review and recommendations. *Milbank Quarterly, 82*, 581-629. doi:10.1111/j.0887-378X.2004.00325.x

Gresham, F. M. (1989). Assessment of treatment integrity in school consultation and pre-referral intervention. *School Psychology Review, 18*, 37-50.

Guba, E. G., & Lincoln, Y. (1982). Epistemological and methodological bases of naturalistic inquiry. *Educational Communications and Technology Journal, 31*, 232-252.

Hall, G., & Loucks, S. (1977). A developmental model for determining whether the treatment is actually implemented. *American Educational Research Journal, 14*, 263-276. doi:10.3102/00028312014003263

Hamre, B., Justice, L., Pianta, R., Kilday, C., Sweeney, B., Downer, J., & Leach, A. (2010). Implementation fidelity of MyTeachingPartner literacy and language activities: Association with preschoolers' language and literacy growth. *Early Childhood Research Quarterly, 25*, 329-347. doi:10.1016/j.ecresq.2009.07.002

Handford, C., Davids, K., Bennett, S., & Button, C. (1997). Skill acquisition in sport: Some applications of an evolving practice ecology. *Journal of Sports Sciences, 15*, 621-640. doi:10.1080/026404197367056

Hansen, W., & McNeal, R. (1999). Drug education practice: Results of an observational study. *Health Education Research, 14*, 85-97. doi:10.1093/her/14.1.85

Harrell, J. S., Gansky, S. A., McMurray, R. G., Bangdiwala, S. I., Frauman, A. C., & Bradley, C.B. (1998). School-based interventions improve heart health in children with multiple cardiovascular disease risk factors. *Pediatrics, 102*, 371-380. Retrieved from http://pediatrics.aappublications.org/

Harrell, J. S., McMurray, R. G., Bangdiwala, S. I., Frauman, A. C., Gansky, S. A., & Bradley, C. B. (1996). Effects of a school-based intervention to reduce cardiovascular disease risk factors in elementary-school children: The Cardiovascular Health in Children (CHIC) study. *Journal of Pediatrics, 128*, 797-805. Retrieved from http://www.jpeds.com/

Heylighen, F., Cilliers, P., & Gershenson, C. (2007). Complexity and philosophy. In J. Bogg & R. Geyer (Eds.), *Complexity, Science and Society* (pp. 117-132). Oxford, UK: Radcliffe.

Hodge, S. R., Ammah, J. O. A., Casebolt, K., Lamaster, K., & O'Sullivan, M. (2004). High school general physical education teachers' behaviors and beliefs associated with inclusion. *Sport, Education and Society, 9*(3), 395-419. doi:10.1080/1357332042331302458

Hoerr, S. L., Nelson, R. A., Essex-Sorlie, D. (1988). Treatment and follow-up of obesity in adolescent girls. *Journal of Adolescent Health Care, 9,* 28-37. Retrieved from http://www.sciencedirect.com/science/journal/01970070

Hohmann, A., & Shear, M. (2002). Community-based intervention research: Coping with the "noise" of real life in study design. *American Journal of Psychiatry, 159*, 201-207. Retrieved from http://ajp.psychiatryonline.org

Huang, F., & Moon, T. (2009). Is experience the best teacher? A multilevel analysis of teacher characteristics and student achievement in low performing schools. *Educational Assessment, Evaluation and Accountability, 21*, 209-234. doi:10.1007/s11092-009-9074-2

Jamner, M. S., Spruijt-Metz, D., Bassin S., & Cooper, D. M. (2004). A controlled evaluation of a school-based intervention to promote physical activity among sedentary adolescent females: Project FAB. *Journal of Adolescent Health, 34,* 279-289. doi:10.1016/j.jadohealth.2003.06.003

Jess, M., Atencio, M., & Thorburn, M. (2011). Complexity theory: Supporting curriculum and pedagogy developments in Scottish physical education. *Sport, Education, and Society, 16*, 179-199. doi:10.1080/13573322.2011.540424

John, D., & Freedson, P. (2012). Actigraph and actical physical activity monitors: A peek under the hood. *Medicine & Science in Sports & Exercise, 44*(Suppl. 1), S86-89. doi:10.1249/MSS.0b013e3182399f5e

Kain, J., Uauy, R., Albala, Vio F., Cerda, R., & Leyton, B. (2004). School-based obesity prevention in Chilean primary school children: Methodology and evaluation of a

controlled study. *International Journal of Obesity, 28,* 483-493. doi:10.1038/sj.ijo.0802611

Kam, C., Greenberg, M. T., & Walls, C. T. (2003). Examining the role of implementation quality in school-based prevention using PATHS curriculum. *Prevention Science, 4*, 55-63. doi:10.1023/A:1021786811186

Katz, L. G. (1972). Developmental stages of preschool teachers. *Elementary School Journal, 73*(1), 50-54. doi:10.1086/460731

Kauffman, S. A. (1995). *At home in the universe: The search for the laws of self organization and complexity.* Oxford, UK: Oxford University Press.

Kelder, S. H., Perry, C. L., & Klepp, K. I. (1993). Community-wide youth exercise promotion: Long-term outcomes of the Minnesota Heart Health Program and the Class of 1989 Study. *Journal of School Health, 63,* 218-223. Retrieved from http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1746-1561

Kellert, S. (1993). *In the wake of chaos: Unpredictable order in dynamical systems.* Chicago, IL: University of Chicago Press.

Kent, A. M. (2004). Improving teacher quality through professional development. *Education, 124*, 427-435. Retrieved from http://www.projectinnovation.biz/education_2006.html

Klinger, J., Ahwee, S., Pilonieta, P., & Mendez, R. (2003). Barriers and facilitators in scaling up research-based practices. *Exceptional Children, 69*, 411-429. Retrieved from http://journals.cec.sped.org/ec/

Klinger, J., Cramer, E., & Harry, B. (2006). Challenges in the implementation of Success for All in four high-needs urban schools. *The Elementary School Journal, 106*, 333-349. doi:10.1086/503635

Kramer, L., Laumann, G., & Brunson, L. (2000). Implementation and diffusion of the Rainbows program in rural communities: Implications for school-based prevention programming. *Journal of Educational and Psychological Consultation, 11*, 37-64. doi:10.1207/s1532768Xjepc1101_04

Kurki, A., Boyle, A., & Aladjem, D. (2006). Implementation: Measuring and explaining the fidelity of CSR Implementation. *Journal of Education for Students Placed At Risk, 11*, 255-277. doi:10.1207/s10824669espr110304_3

Langston, C. (1990). Computation at the edge of chaos: Phase transition and emergent computation. *Physica D: Nonlinear Phenomena, 42,* 12-37. Retrieved from http://www.journals.elsevier.com/physica-d-nonlinear-phenomena/

Leithwood, K., Day, C., Sammons, P., Harris, A., & Hopkins, D. (2006). *Successful school leadership: What it is and how it influences pupil learning.* Nottingham, England: National College for School Leadership, Department for Education and Skills, University of Nottingham. Retrieved from https://www.education.gov.uk/publications/RSG/_arc_SOP/Page6/RR800

Light, R. (2008). Complex learning theory – Its epistemology and its assumptions about learning: Implications for physical education. *Journal of Teaching in Physical Education, 27,* 21-37. Retrieved from http://journals.humankinetics.com/jtpe

Light, R., & Fawns, R. (2003). Knowing the game: Integrating speech and action through TGfU. *Quest, 55,* 161-176. doi:10.1080/00336297.2003.10491797

Lillehoj, C. J., Griffin, K. W., & Spoth, R. (2004). Program provider and observer ratings of school-based preventive intervention implementation: Agreement and relation to youth outcomes. *Health Education & Behavior, 31*, 242-257. doi:10.1177/1090198103260514

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry.* Newbury Park, CA: Sage.

Lipsey, M. W., & Cordray, D. S. (2000). Evaluation methods for social intervention. *Annual Review of Psychology, 51*, 345-375. Retrieved from http://www.annualreviews.org/journal/psych

Locke, L. F. (1992) Changing secondary school physical education. *Quest, 44*, 361-372.

Luepker, R. V., Perry, C. L., McKinlay, S. M., Nader, P. R., Parcel, G. S., Stone, E. J., . . . Johnson, C. C. (1996). Outcomes of a field trial to improve children's dietary patterns and physical activity: The Child and Adolescent Trial for Cardiovascular Health (CATCH). *Journal of the American Medical Association, 275*, 768-776. Retrieved from http://jama.jamanetwork.com/journal.aspx

Lund, J. L. (1993). Same sneakers, new tricks: Curricular alternatives for physical education. *National Association of Secondary School Principals Bulletin, 77,* 97-100. doi:10.1177/019263659307755613

Lynch, S. (2007, April). *A model for fidelity of implementation in a study of a science curriculum unit: Evaluation based on program theory.* In C. L. O'Donnell (Chair), *Analyzing the relationship between fidelity of implementation (FOI) and*

student outcomes in a quasi-experiment. Symposium conducted at the Annual Meeting of the American Educational Research Association, Chicago, IL. Retrieved from http://www.gwu.edu/~scale-up/cp2007.html

Lynch, S., & O'Donnell, C. (2005, April). *The evolving definition, measurement, and conceptualization of fidelity of implementation in scale-up of highly rated science curriculum units in diverse middle schools.* In S. Lynch (Chair) & C. L. O'Donnell, *Fidelity of implementation in implementation and scale-up research designs: Applications from four studies of innovative science curriculum materials and diverse populations.* Symposium conducted at the Annual Meeting of the American Educational Research Association, Montreal, Canada. Retrieved from http://www.gwu.edu/~scale-up/fidelity.html

Manios, Y., Moschandreas, J., Hatzis, C., & Kafatos, A. (1999). Evaluation of a health and nutrition education program in primary school children of Crete over a three-year period. *Preventive Medicine, 28,* 149-159. Retrieved from http://www.journals.elsevier.com/preventive-medicine/

Marion, R. (1999). *The edge of organization: Chaos and complexity theories of formal social systems.* New York, NY: Corwin Press.

Martens, B., & McIntyre, L. (2009). The importance of treatment integrity in school-based behavioral intervention. In A. Akin-Little (Ed.), *Behavioral interventions in schools: Evidence-based positive strategies* (pp. 59-71). Washington, DC: American Psychological Association.

McGrew, J., Bond, G., Dietzen, L., & Salyers, M. (1994). Measuring the fidelity of implementation of a mental health program model. *Journal of Consulting and Clinical Psychology, 62*, 670-678. Retrieved from http://www.apa.org/pubs/journals/ccp/index.aspx

McKenzie, T. L., Catellier, D. J., Conway, T., Lytle, L. A., Grieser, M., Webber, L. A., . . . Elder, J. P. (2006). Girls' activity levels and lesson contexts in middle school PE: TAAG baseline. *Medicine & Science in Sports & Exercise, 38*, 1229-1235. doi:10.1249/01.mss.0000227307.34149.f3

McLeod, B. D., & Southam-Gerow, M. A. (2009). Conceptual and methodological issues in treatment integrity measurement. *School Psychology Review, 28*, 541-546. Retrieved from http://www.nasponline.org/publications/spr/spr384index.aspx

McMillan, E. (2008). *Complexity, management and the dynamics of change.* London, UK: Routledge.

McMurray, R. G., Harrell, J. S., Bangdiwala, S. I., Bradley, C. B., Deng, S., & Levine, A. (2002). A school-based intervention can reduce body fat and blood pressure in young adolescents. *Journal of Adolescent Health, 31,* 125-132. Retrieved from http://www.jahonline.org/

Melanson, E. L. Jr., & Freedson P. S. (1995). Validity of the computer science and applications, Inc. (CSA) activity monitor. *Medicine & Science in Sports & Exercise, 27*, 934-940. Retrieved from http://www.ovid.com/webapp/wcs/stores/servlet/ProductDisplay?storeId=13051&catalogId=13151&langId=-1&partNumber=Prod-586

Mertens, D. M. (2003). Mixed methods and the politics of human research: The transformative-emancipatory perspective. In A. Tashakkori & C. Teddlie (Eds.), *Handbook on mixed methods in the behavioral and social sciences* (pp. 135-164). Thousand Oaks, CA: Sage Publications.

Meyer, J. W., & Rowan, B. (1983). Institutionalized organizations: Formal structure as myth and ceremony. In J. W. Meyer & W. R. Scott (Eds.), *Organizational environments: Ritual and rationality* (pp. 21-44). Thousand Oaks, CA: Sage.

Mills, S., & Ragan, T. (2000). A tool for analyzing implementation fidelity of an integrated learning system. *Educational Technology Research & Development, 48*(4), 21-41. Retrieved from http://www.springer.com/education+%26+language/learning+%26+instruction/journal/11423

Moncher, F., & Prinz, R. (1991). Treatment fidelity in outcome studies. *Clinical Psychology Review, 11*, 247-266. Retrieved from http://www.sciencedirect.com/science/journal/02727358

Morrison, K. (2006, November). *Complexity theory and education.* Paper presented at the Asian-Pacific Education Research Association Conference, Macau, Hong Kong. Retrieved from http://edisdat.ied.edu.hk/pubarch/b15907314/full_paper/SYMPO-000004_Keith%20Morrison.pdf

Morrison, K. (2008). Educational philosophy and the challenge of complexity theory. *Educational Philosophy and Theory, 40*, 19-34. doi:10.1111/j.1469-5812.2007.00394.x

Mowbray, C. T., Holter, M. C., Teague, G. B., & Bybee, D. (2003). Fidelity criteria: Development, measurement, and validation. *American Journal of Evaluation, 24*, 315-340. doi:10.1177/109821400302400303

Nader, P. R., Stone, E. J., Lytle, L. A., Perry, C. L., Osganian, S. K., Kelder, S., . . . Luepker, R. V. (1999). Three-year maintenance of improved diet and physical activity: The CATCH cohort. Child and Adolescent Trial for Cardiovascular Health. *Archives of Pediatrics & Adolescent Medicine, 153,* 695-704. Retrieved from http://archpedi.jamanetwork.com/journal.aspx

National Center for Education Statistics. (2012). *120 Years of Literacy.* Washington, DC: Institute of Education Sciences. Retrieved from http://nces.ed.gov/naal/lit_history.asp

National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups* (NIH Publication No. 00-4754). Washington, DC: Government Printing Office. Retrieved from http://www.nichd.nih.gov/publications/pubs/nrp/Documents/report.pdf

National Institutes of Health. (2011). *Dissemination and implementation.* Retrieved from http://obssr.od.nih.gov/scientific_areas/translation/dissemination_and_implementation/index.aspx

National Research Council Committee for a Review of the Evaluation Data on the Effectiveness of NSF-Supported and Commercially Generated Mathematics

Curriculum Materials, Mathematical Sciences, Education Board, Center for Education, Division of Behavioral and Social Sciences and Education. (2004). *On evaluating curricular effectiveness: Judging the quality of K-12 mathematics evaluations*. Washington, DC: National Academies Press.

Neumark-Sztainer, D., Story, M., Hannan, P. J., & Rex, J. (2003). New Moves: A school-based obesity prevention program for adolescent girls. *Preventive Medicine, 37,* 41-51. doi:10.1016/S0091-7435(03)00057-4

No Child Left Behind (NCLB) Act of 2001, 20 U.S.C.A. § 6301 *et seq.* (West 2003)

O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K-12 curriculum intervention research. *Review of Educational Research, 78*, 33-84. doi:10.3102/0034654307313793

Ogden, C. L., Carroll, M. D., Kit, B. K. & Flegal, K. M. (2012). Prevalence of obesity and trends in body mass index among US children and adolescents, 1999-2010. *Journal of the American Medical Association, 307*, 483-90. doi:10.1001/jama.2012.40

O'Hare, T. (2005). *Evidence-based practices for social workers: An interdisciplinary approach*. Chicago, IL: Lyceum Books.

Osberg, D., Biesta, G. & Cilliers, P. (2008) From representation to emergence: Complexity's challenge to the epistemology of schooling. *Educational Philosophy and Theory, 40*, 213-227. doi:10.1111/j.1469-5812.2007.00407.x

Owens, S., Gutin, B., Allison, J., Riggs, S., Ferguson, M., Litaker, M., & Thompson, W. (1999). Effect of physical training on total and visceral fat in obese children. *Medicine & Science in Sports & Exercise, 31,* 143-148. doi:10.1097/00005768-199901000-00022

Palardy, G., & Rumberger, R. (2008). Teacher effectiveness in first grade: The importance of background qualifications, attitudes, and instructional practices for student learning. *Educational Evaluation and Policy Analysis, 30*, 111-140. doi:10.3102/0162373708317680

Parcel, G. S., Ross, J. G., Lavin, A. T., Portnoy, B., Nelson, G. D., & Winters, F. (1991). Enhancing implementation of the teenage health teaching modules. *Journal of School Health, 61*, 35-38. doi:10.1111/j.1746-1561.1991.tb07857.x

Parker, M. B., & Curtner-Smith, M. (2005). Health-related fitness in sport education and multi-activity teaching. *Physical Education and Sport Pedagogy, 10*, 1-18. doi:10.1080/1740898042000334872

Pate, R. R., Saunders, R. P., Ward, D. S., Felton, G., Trost, S. G., & Dowda, M. (2003). Evaluation of a community-based intervention to promote physical activity in youth: Lessons from Active Winners. *American Journal of Health Promotion, 17,* 171-182. doi:10.4278/0890-1171-17.3.171

Pate, R. R., Ward, D. S., Saunders, R. P., Felton, G., Dishman, R. K., & Dowda, M. (2005). Promotion of physical activity among high-school girls: A randomized controlled trial. *American Journal of Public Health, 95*, 1582-1587. doi:10.2105/AJPH.2004.045807

Patton, M. Q. (1997). Implementation evaluation: What happened in the program? In M.Q. Patton (Ed.), *Utilization-focused evaluation: The new century text* (pp. 195-211). Beverly Hills, CA: Sage.

Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage.

Paulhus, D., & Vazire, S. (2007). The self-report method. In R. Robbins, C. Fraley, & R.Krueger (Eds.), *Handbook of research methods in personality* (pp. 224-239). NewYork, NY: Guildford Press.

Peat, F. D. (2002). *From certainty to uncertainty: The story of science and ideas in the twentieth century.* Washington, DC: Joseph Henry Press.

Penuel, W. R., & Means, B. (2004). Implementation variation and fidelity in an inquiry science program: An analysis of GLOBE data reporting patterns. *Journal of Research in Science Teaching, 41*, 294-315. doi:10.1002/tea.20002

Public Schools of North Carolina, State Board of Education, Department of Public Instruction. (2011). *North Carolina essential standards: Healthful living.* Retrieved from http://www.ncpublicschools.org/acre/standards/new-standards/#healthful

Qualtrics Labs, Inc. (Version 31,640) [Computer Software]. (2012). Provo, UT: Qualtrics Labs, Inc. Available from http://www.qualtrics.com

Resnick, B., Bellg, A., Borrelli, B., DeFrancesco, C., Breger, R., Hecht, J., Sharp, D.,… Czajkowski, S. (2005). Examples of implementation and evaluation of treatment fidelity in the BCC studies: Where we are and where we need to go. *Annals of*

Behavioral Medicine, 29, 46-54. Retrieved from http://www.springer.com/medicine/journal/12160

Resnicow, K., Davis, M., Smith, M., Lazarus-Yaroch, A., Baranowski, T., &Baranowski, J. (1998). How best to measure implementation of school health curricula: A comparison of three measures. *Health Education Research, 13*, 239-250. doi:10.1093/her/13.2.239

Richardson, K. & Cilliers, P. (2001). Special editor's introduction: What is complexity science? A view from different directions. *Emergence: Complexity and Organization, 3*(1), 5-23. Retrieved from http://emergentpublications.com/ECO/about_eco.aspx?AspxAutoDetectCookieSupport=1

Riley, B., Taylor, M., & Elliott, S. (2003). Organizational capacity and implementation change: A comparative case study of heart health promotion in Ontario public health agencies. *Health Education Research, 18*, 754-769. doi:10.1093/her/cyf051

Ringwalt, C. L., Ennett, S., Johnson, R., Rohrbach, L. A., Simons-Rudolph, A., Vincus, A., & Thorne, J. (2003). Factors associated with fidelity to substance use prevention curriculum guides in the nation's middle schools. *Health Education & Behavior, 30*, 375-391. doi:10.1177/1090198103030003010

Rivkin, S., Hanushek, E., & Kain, J. (2005). Teachers, schools, and academic achievement. *Econometrica, 73*, 417-458. Retrieved from http://www.jstor.org/stable/3598793

Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). New York, NY: Free Press.

Rohrbach, L. A., Graham, J. W., & Hansen, W. B. (1993). Diffusion of a school-based substance abuse prevention program: Predictors of program implementation. *Preventive Medicine, 22*, 237-260. doi:10.1006/pmed.1993.1020

Ross, J., McDougall, D., Hogaboam-Gray, A., & LeSage, A. (2003). A survey measuring elementary teachers' implementation of standards-based mathematics teaching. *Journal for Research in Mathematics, 34*, 344-363. Retrieved from http://www.jstor.org/stable/30034787

Ruiz-Primo, M. A. (2006). A multi-method and multi-source approach for studying fidelity of implementation. CSE Report 677 for the National Center for Research on Evaluation, Standards, and Student Testing. Los Angeles, CA. Retrieved from http://www.eric.ed.gov/PDFS/ED492864.pdf

Sallis, J. F., McKenzie, T. L., Alcaraz, J. E., Kolody, B., Faucette, N., & Hovell, M. F. (1997). The effects of a 2-year physical education program (SPARK) on physical activity and fitness in elementary school students. Sports, Play, and Active Recreation for Kids. *American Journal of Public Health, 87,* 1328-1334. Retrieved from http://ajph.aphapublications.org/

Sallis, J. F., McKenzie, T. L., Beets, M. W., Beighle, A., Erwin, H., & Lee, S. (2012). Physical education's role in public health: Steps forward and backward over 20 years and HOPE for the future. *Research Quarterly for Exercise and Sport, 83*, 125-135. doi:10.1080/02701367.2012.10599842

Sallis, J. F., McKenzie, T. L., Conway, T. L., Elder, J. P., Prochaska, J. J., Brown, M., . . . Alcaraz, J. E. (2003). Environmental interventions for eating and physical

activity: A randomized controlled trial in middle schools. *American Journal of Preventive Medicine, 24,* 209-217. doi:10.1016/S0749-3797(02)00646-3

Sanchez, V., Steckler, A., Nitirat, P., Hallfors, H., Cho, H., & Brodish, P. (2007). Fidelity of implementation in a treatment effectiveness trial of Reconnecting Youth. *Health Education Research, 22*, 95-107. doi:10.1093/her/cyl052

Sardar, Z. & Abrams, I. (2008). *Introducing chaos: A graphic guide.* Cambridge, UK: Icon Books.

Sato, T., Hodge, S. R., Murata, N. M., & Maeda, J. K. (2007). Japanese physical education teachers' beliefs about teaching students with disabilities. *Sport, Education and Society, 12*(2), 211-230. doi:10.1080/13573320701287536

Saunders, R. P., Ward, D., Felton, G. W., Dowda, M., & Pate, R. R. (2006). Examining the link between program implementation and behavior outcomes in the lifestyle education for activity program (LEAP). *Evaluation and Program Planning, 29*, 352-364. doi:10.1016/j.evalprogplan.2006.08.006

Schein, E. H. (2010). *Organizational culture and leadership* (4th ed.). San Francisco, CA: Jossey-Bass.

Seltz, J. (2008, May). Looking back, looking forward: A focus on high school reform. *Educational Leadership, 65*(8), 96. Retrieved from http://www.ascd.org/publications/educational-leadership/may08/vol65/num08/Looking-Back,-Looking-Forward@-A-Focus-on-High-School-Reform.aspx

Siedentop, D., Mand, C., & Taggart, A. (1986) Physical education: Teaching and curriculum strategies for grades 5-12. Palo Alto, CA: Mayfield Publishing.

Siegel, D. J. (2001). Toward an interpersonal neurobiology of the developing mind: Attachment relationships, "mindsight," and neural integration. *Infant Mental Health Journal, 22*, 67-94. doi:10.1002/1097-0355(200101/04)22:1<67::AID-IMHJ3>3.0.CO;2-G

Simons-Morton, B. G., Parcel, G. S., Baranowski, T., Forthofer, R., & O'Hara, N. M. (1991). Promoting physical activity and a healthful diet among children: Results of a school-based intervention study. *American Journal of Public Health, 81,* 986-991. Retrieved from http://ajph.aphapublications.org/

Slavin, R. (2003). A reader's guide to scientifically based research: Learning how to assess the validity of education research is vital for creating effective, sustained reform. *Educational Leadership, 60*(5), 12-16. Retrieved from http://www.ascd.org/publications/educational-leadership.aspx

Smith, S. W., Daunic, A. P., & Taylor, G. G. (2007). Treatment fidelity in applied educational research: Expanding the adoption and application of measures to ensure evidence-based practice. *Education and Treatment of Children, 30*(4), 121-134. doi:10.1353/etc.2007.0033

Smylie, M. A. (1988). The enhancement function of staff development: Organizational and psychological antecedents to individual teacher change. *American Educational Research Journal, 25*, 1-30. doi:10.3102/00028312025001001

Sobeck, J. L. (2006). Lessons learned from implementing school-based substance abuse prevention curriculums. *Children & Schools, 28*(2), 77-85. doi:10.1093/cs/28.2.77

Stacey, R. D. (2003). *Strategic management and organisational dynamics. The challenge of complexity* (4th ed.). Harlow, UK: Pearson Education/Prentice Hall.

Stead, M., Stradling, R., Macneil, M., Mackintosh, A., & Minty, S. (2007). Implementation evaluation of the blueprint multi-component drug prevention programme: Fidelity of school component delivery. *Drug and Alcohol Review, 26*, 653-664. doi:10.1080/0959230701613809

Stein, M. L., Berends, M., Fuchs, D., McMaster, K., Sáenz, L., Yen, L., . . . Compton, D. L. (2008). Scaling up an early reading program: Relationships among teacher support, fidelity of implementation, and student performance across different sites and years. *Educational Evaluation and Policy Analysis, 30*, 368-388. doi:10.3102/0162373708322738

Stevenson, K. R. (2010). Educational trends shaping school planning, design, construction, funding and operation. *National Clearinghouse for Educational Facilities.* Washington, DC: National Institute of Building Sciences. Retrieved from http://www.ncef.org/pubs/educationaltrends.pdf

Storey, B. & Butler, J. (2012). Complexity thinking in PE: Game-centered approaches, games as complex adaptive systems, and ecological values. *Physical Education and Sport Pedagogy,* 1-17. doi:10.1080/17408989.2011.649721

Stroot, S. A., & Whipple, C. E. (2003). Organizational socialization: Factors affecting beginning teachers. In S. J. Silverman & C. D. Ennis (Eds.), *Student learning in*

physical education: Applying research to enhance instruction (pp. 311-328). Champaign, IL: Human Kinetics.

Sun, H., Chen, A., Zhu, X., & Ennis, C.D. (2012). Curriculum matters: Learning science-based fitness knowledge in constructivist physical education. *The Elementary School Journal, 113*, 215-229. doi:10.1086/667405

Szulanski, G., & Winter, S. (2002). Getting it right the second time. *Harvard Business Review, 80*, 62-69. Retrieved from https://hbdm.hbsp.harvard.edu/hbr/sub/subscribe5.html#form

Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches.* Thousand Oaks, CA: Sage.

Tashakkori, A., & Teddlie, C. (2010). *Handbook of mixed methods in social and behavioral research* (2nd ed.). Thousand Oaks, CA: Sage.

Teague, G., Drake, R., & Ackerson, T. (1995). Evaluating use of continuous treatment teams for persons with mental illness and substance abuse. *Psychiatric Services, 46*, 689-695. Retrieved from http://ps.psychiatryonline.org/journal.aspx?journalid=18

Teague, G., Ganju, V., Hornik, J., Johnson, J., & McKinney, J. (1997). The MHSIP mental health report card: A consumer-oriented approach to monitoring the quality of mental health plans. *Evaluation Review, 21*, 330-341. doi:10.1177/0193841X9702100307

Thelen, E. & Smith, L.B. (1994). *A dynamic systems approach to the development of cognition and action.* Cambridge, MA: MIT Press.

Tucker, A. R., & Blythe, B. (2008). Attention to treatment fidelity in social work outcomes: A review of the literature from the 1990s. *Social Work Research, 32*(3), 185-190. doi:10.1093/swr/32.3.185

Tyack, D., & Cuban, L. (1995). *Tinkering toward utopia: A century of public school reform.* Cambridge, MA: Harvard University Press.

U.S. Department of Education. (2002, August). *Guidance on the comprehensive school reform program.* Washington, DC: Author. Retrieved from http://www2.ed.gov/programs/compreform/guidance/index.html

U.S. Department of Education. (2011). *What Works Clearinghouse procedures and standards handbook* (Version 2.1). Washington, DC: Institute of Education Sciences. Retrieved from http://ies.ed.gov/ncee/wwc/documentsum.aspx?sid=19

U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Division of Adolescent and School Health. (2010). *Strategies to improve the quality of physical education.* Retrieved from http://www.cdc.gov/healthyyouth/physicalactivity/pdf/quality_pe.pdf

Vandongen, R., Jenner, D. A., Thompson, C., Taggart, A. C., Spickett, E. E., Burke, V., . . . Dunbar, D. L. (1995). A controlled evaluation of a fitness and nutrition intervention program on cardiovascular health in 10- to 12-year-old children. *Preventive Medicine, 24*, 9-22. Retrieved from http://www.journals.elsevier.com/preventive-medicine/

Wanless, S. B., Patton, C. L., Rimm-Kaufman, S. E., & Deutsch, N. L. (2013). Setting-level influences on implementation of the Responsive Classroom approach. *Prevention Science, 14*, 40-51. doi:10.1007/s11121-012-0294-1

Wayne, A., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational Research, 73*, 89-122. doi:10.3102/00346543073001089

Webber, L. S., Catellier, D. J., Lytle, L. A., Murray, D. M., Pratt, C. A., Young, D. R., . . . Pate, R. R. (2008). Promoting physical activity in middle school girls: Trial of Activity for Adolescent Girls. *American Journal of Preventive Medicine, 34*, 173-184. doi:10.1016/j.amepre.2007.11.018

Wolf, M. M. (1978). Social validity: The case for subjective measurement or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis, 11*(2), 203-214. doi:10.1901/jaba.1978.11-203

Young, R., DeMoor, C., Wildey, M., Gully, S., Hovell, M., & Elder, J. (1990). Correlates of health facilitator performance in a tobacco use prevention program: Implication for recruitment. *Journal of School Health, 60*, 463-467. doi:10.1111/j.1746-1561.1990.tb05979.x

Ysseldyke, J. R., Spicuzza, S., Kosciolek, S., & Boys, C. (2003). Effects of a learning information system on mathematics achievement and classroom structure. *Journal of Educational Research, 96*, 163-173. doi:10.1080/00220670309598804

APPENDIX A

SHL IRB APPROVAL PAGE

THE UNIVERSITY of NORTH CAROLINA
GREENSBORO

OFFICE OF RESEARCH INTEGRITY
2718 Beverly Cooper Moore and Irene Mitchell Moore
Humanities and Research Administration Bldg.
PO Box 26170
Greensboro, NC 27402-6170
336.256.1482
Web site: www.uncg.edu/orc
Federalwide Assurance (FWA) #216

To: Catherine Ennis
Dept Of Kinesiology
250 HHP Building

From: UNCG IRB

Authorized signature on behalf of IRB

Approval Date: 1/11/2013
Expiration Date of Approval: 1/10/2014

RE: Notice of IRB Approval by Expedited Review (under 45 CFR 46.110)
Submission Type: Renewal
Expedited Category: 7.Surveys/interviews/focus groups,4.Noninvasive clinical data
Study #: 11-0009
Sponsors: US DHHS National Institutes of Health (NIH)
Study Title: The Science of Healthful Living

This submission has been approved by the IRB for the period indicated.

Study Description:

This project will design and field test a science-enriched middle school healthful living curriculum to increase students' knowledge and interest in health related science, enhance their intention to pursue a life science related career, and improve the communities' understandings of NIH funded clinical and basic research.

Submission Description:

Renewal request, dated 1/9/13. Enrollment of new participants continues.

Regulatory and other findings:

This research, which involves children, meets criteria at 45 CFR 46.404 (research involving no greater than minimal risk). Permission of one parent or guardian is sufficient.

This research meets criteria for a waiver of consent entirely according to 45 CFR 46.116(d).(Parents)

Investigator's Responsibilities

Federal regulations require that all research be reviewed at least annually. It is the Principal Investigator's responsibility to submit for renewal and obtain approval before the expiration date.

You may not continue any research activity beyond the expiration date without IRB approval. Failure to receive approval for continuation before the expiration date will result in automatic termination of the approval for this study on the expiration date.

Signed letters, along with stamped copies of consent forms and other recruitment materials will be scanned to you in a separate email. These consent forms must be used unless the IRB hasgiven you approval to waive this requirement.

You are required to obtain IRB approval for any changes to any aspect of this study before they can be implemented (use the modification application available at http://www.uncg.edu/orc/irb.htm). Should any adverse event or unanticipated problem involving risks to subjects or others occur it must be reported immediately to the IRB using the "Unanticipated Problem/Event" form at the same website.

CC:
Ang Chen, Dept Of Kinesiology
OSP, (Office of Sponsored Programs), Non-IRB Review Contact
William Walters, (Contracts and Grants), Non-IRB Review Contact
ORC, (ORC), Non-IRB Review Contact

APPENDIX B

CONSENT AND ASSENT FORMS

RESEARCH INFORMATION AND CONSENT FORM
(Teachers – Experimental Group, teaching the Healthful Living Curriculum)

To be completed by non-student participant or student participant aged 18 years and above.

Research Project Name: The Science of Healthful Living

Sponsoring Organization: National Institutes of Health and University of North Carolina at Greensboro

Principal Researcher: Dr. Catherine D. Ennis Telephone: 256-8565; Email: c_ennis@uncg.edu

The purpose of this research is to test a new *Science of Healthful Living* curriculum for middle school students. The curriculum is consistent with the North Carolina Healthful Living Standard Course of Study for physical education and health education. Middle school physical education teachers will be trained to teach the curriculum and will provide feedback concerning their evaluation of the curriculum. The *Science of Healthful Living* curriculum consists of two, 20-lesson units for 6th, 7th and 8th grades students developed by master teachers in the Piedmont Triad School districts. The unit consists of learning activities, knowledge tests, physical activities, student journals, and student satisfaction surveys.

To evaluate the curriculum, we will observe the 6th through 8th grade classes in physical education, review students' written work in physical education, monitor students' physical activity levels, and interview 6 students in each grade who demonstrate a range of knowledge and understandings about the Healthful Living content. We will ask you to complete questionnaires regarding your satisfaction with the curriculum, evaluate the Ease of Using the curriculum, and make suggestions and recommendations to improve the teacher training and the curriculum.

We will ask you to participate in individual interviews or focus groups to explain your experience when planning and teaching the curriculum and to make suggestions or recommendations for revisions and improvements to the process. Interviews and focus groups typically occur at the conclusion of the units and last for about 60 min. They will be conducted at a time and place of your convenience. Your answers will be combined with answers from others and may be discussed (without names or other identifiers) with other physical education teachers at staff development sessions and communicated to other teachers through presentations and publications. Your name will not be disclosed without your permission.

Specifically, in this research we will:

(a) Ask you to provide general information about yourself, such as where you received your bachelors degree, if you have a masters degree, years of teaching experience, ethnicity, and gender.
(b) Ask you to attend 18 hrs. of teacher training to teach the *Science of Healthful Living* curriculum each year during years 1-3 of the project
(c) Ask you to complete the Teacher Ease of Use online inventory following each unit with your recommendations for curriculum and training improvements.
(d) Observe lessons that you teach to 6th, 7th, & 8th grade students.
(e) Ask questions when needed for clarification and to increase our understanding of the curriculum
(f) Select 6 students in each grade to interview based on their worksheet responses.
(g) Select 6 students in each grade to wear accelerometers to measure lesson intensity
(h) Interview you following the units to gather your thoughts, comments, and suggestions for teaching fitness concepts. The interview will be tape-recorded. Because your voice will be potentially identifiable by anyone who hears the tape, confidentiality for things said on the tape cannot be guaranteed, although the Dr. Ennis will try to limit access to the tape as described below.

Time Required for this project is approximately 18 hours each year in training to teach the curriculum; approximately 40 hrs. preparing to teach lessons each year; approximately 2 hrs /yr completing surveys and questionnaires; & approximately 1 hr in interviews each year.

UNCG IRB
Approved Consent Form
Valid 1/11/13 to 1/10/14

None of the procedures used in this research is experimental. The Institutional Review Board at the University of North Carolina at Greensboro has determined that participation in this study poses no risk to you. There are no costs to you or payments to you as a result of participation in this study. There are no immediate benefits to you, your students, or to society. We hope that the information about effective teaching and student learning will be useful to you and to other teachers in future teacher trainings and will assist in increasing student learning in physical education.

If you have any concerns about your rights, how you are being treated or if you have questions, want more information or have suggestions, please contact Research Compliance Office at UNCG at (336) 256-1482.

If you have any questions or concerns about this research project, you may contact Dr. Ennis at (336) 256-8565; email: c_ennis@uncg.edu; Additional information about Dr. Ennis' research is available at the following Website: http://www.uncg.edu/ess/faculty/cennis.html.

Participants' Rights and Assurances

I understand that during the course of this project my responses will be kept strictly confidential and that none of the data released in this study will identify me by name or any other identifiable data, descriptions or characterizations unless disclosure is required by law. Your name will not be used on the audio-tape or on any materials associated with this project. Copies of your interview tape(s) and transcripts will be kept in the locked file in Dr. Ennis' office and will be destroyed following the completion of the research.

Furthermore, I understand that I may discontinue my participation in this project at any time or refuse to respond to any questions to which I choose not to respond. I am a voluntary participant and have no liability or responsibility for the implementation, methodology, claims, substance or outcomes resulting from this research project. I also am aware that my decision not to participate will not result in any adverse consequences or disparate treatment due to that decision. If know that if I choose to withdraw, I may request that any identifiable data that has been collected be destroyed.

I fully understand that this research is being conducted for constructive educational purposes and that my signature gives my consent to voluntarily participate in this project. I have received information about the aforementioned research project. Having thoroughly read and reviewed the application, I am familiar with the purpose, methods, scope, and intent of the research project.

School Name: _____

Your Name: _____Position: Health and Physical Education Teacher

Home Address_____ _____Cell phone_____

Please circle your response: **I am willing/not willing to participate in the research project.**

Your Signature_____ Date_____

Please Print Your Name _____

UNCG IRB
Approved Consent Form
Valid 1/11/13 to 1/10/14

GUARDIAN/PARENT RESEARCH INFORMATION AND CONSENT FORM
(STUDENTS UNDER 18 – Selected to wear Accelerometers & Participate in Interviews/ Focus Groups)

To be completed by the parent/legal guardian and school aged participant under 18 years.

UNCG IRB
Approved Consent Form
Valid 1/11/13 to 1/10/14

Research Project Name: The Science of Healthful Living

Sponsoring Organization: National Institutes of Health; University of North Carolina at Greensboro

Principal Researchers: Dr. Catherine Ennis (UNCG) 256-8565

The purpose of this research is to test a new Healthful Living Curriculum for middle school students. The curriculum is consistent with the North Carolina Healthful Living Standard Course of Study for physical education and health education. The Healthful Living curriculum consists of two units for 6th, 7th and 8th grades students developed by master teachers in Piedmont Triad School districts. The unit consists of learning activities, knowledge tests, physical activities, worksheets, and student satisfaction surveys.

Two ways to test to determine if the new Healthful Living Curriculum is effective is to (1) ask student what they have learned in the curriculum and (2) measure how much exercise they actually receive when participating in the lessons. To determine what students have learned, we will interview 6 students in each physical education grade who demonstrate a range of knowledge and understandings about the Healthful Living content. Interviews will be conducted individually or in small groups of 4-6 students. Students will be interviewed for approximately 30 min. before the first lesson and after the last lesson in each curriculum unit. Interview questions will be limited to questions about health, physical education, and fitness, for example "Why does your heart beat faster when you exercise?" Additionally, we will ask students about their experiences in the lessons, satisfaction, and motivation to participate in physical activity as a result of the curriculum. Finally we will ask them to make suggestions or recommendations for improvements to the curriculum. To determine how much exercise students receive when participating in the curriculum, we will ask the same 6 students in each physical education class to wear a small "pager" size instrument called an accelerometer. Accelerometers are worn on the waistband of students' shorts and measure the amount of energy and calories expended simply by measuring students' movements.

Specifically, in this research:

1. Six students in each class will be selected for four, 30 min. interviews (at the beginning and end of each unit); conducted individually or in small focus groups. The purpose of the interview is to ask students about what they know (before the unit) or what they learned as a result of the unit.
2. Interviews will be tape recorded to increase accuracy of recording students' answers. Because your child's voice will be potentially identifiable by anyone who hears the tape, confidentiality for things said on the tape cannot be guaranteed, although the Dr. Ennis, the project director will try to limit access to the tape as described below.
3. The same six students will be asked to wear an accelerometer during physical education lessons to measure the number of calories expended in the activities they experience in the lessons. They will receive the accelerometer at the beginning of the lesson and return it at the end. To accurately measure the caloric expenditure, the students' height and weight will be measured to determine their basal metabolic rate. The height/weight measurements will be taken privately in school's nurse room or physical education teachers' office. Results will be reported as an average of students' energy expenditure in that lesson; names and other individual information will never be public.

Time required for this project is 60 min for each unit (four, 30 min. interviews or focus groups). Interviews will be conducted during student's physical education class. Accelerometers will be worn during class time and will not require additional time separate from the regular physical education period.

None of the procedures used in this research is experimental. The Institutional Review Board at the University of North Carolina at Greensboro has determined that participation in this study poses no risk to students. There are no costs to you or payments to you or your child as a result of participation in this study.

There are no immediate benefits to students or to society. We hope that the information about effective teaching, student learning, and students' energy expenditure in these lessons will be useful in future teacher trainings and will assist in increasing student learning.

Participants/Parental Rights and Assurances

As a parent or guardian, I have read the information on page 1 of this information sheet regarding this research. Having read the application, I am familiar with the purpose, methods, scope, and intent of the research project.

I understand that during the course of this project my child's responses will be kept strictly confidential and that none of the data released in this study will identify my child by name or any other identifiable data, descriptions or characterizations unless disclosure is required by law. Students' names will be removed from the interviews and accelerometry data and data will be assigned an identification number unique to this study. Students' interview responses will be matched using their identification number. A master list of student names and identification numbers will be kept in a locked file in the research director, Dr. Ennis', office and will be destroyed after all accelerometry data and interview responses have been matched. Accelerometry data and interview tapes (without names) will be kept in the locked file in Dr. Ennis' office and will be destroyed at the completion of the research.

I understand that my child may refuse to wear the accelerometer or be interviewed or to respond to any questions to which he/she chooses not to answer. My child is a voluntary participant and has no liability or responsibility for the implementation, methodology, claims, substance, or outcomes resulting from this research project. I am also aware that my child's decision not to participate will not result in any adverse consequences or disparate treatment due to that decision.

If you have any concerns about your child's rights, how s/he is being treated or if you have questions, want more information or have suggestions, please contact the Research Compliance Office at UNCG at (336) 256-1482.

If you have any questions or concerns about this research project, you may contact Dr. Ennis at (336) 256-8565; email: c_ennis@uncg.edu; Additional information about Dr. Ennis' research is available at the following Website: http://www.uncg.edu/ess/faculty/cennis.html.

I fully understand that this research is being conducted for constructive educational purposes and that my signature gives consent for my child to voluntarily participate in this project.

Your Child's Name_____

Home Address_____

Your School's Name _____

Classroom Teacher _____

Your Student's Grade (please circle): 6th 7th 8th Age_____

***Please circle your choice*: I am willing / not willing for my child to participate in the research project.**

Parent's Signature_____ Date_____

Print Name _____

PLEASE CIRCLE, SIGN, AND RETURN THIS PAGE TO YOUR PHYSICAL EDUCATION TEACHER

UNCG IRB
Approved Consent Form
Valid 1/11/13 to 1/10/14

Student Assent Form – Accelerometer & Interviews - Healthful Living Schools

My name is _____ My School is _____

I am in the **6th grade** **7th grade** **8th grade** (please circle your grade).

My Physical Education Teacher's Name is _____

My Classroom Teacher's Name is _____

What is this about?
UNCG researchers have worked with middle school physical education teachers to develop a new physical education curriculum. Before the curriculum can be approved it needs to be tested with middle school students to determine if it increases your knowledge, is physically active, and is enjoyable. The UNCG researchers have selected you to wear an accelerometer and to be interviewed or to participate in a small group interview of 4-6 students before and after each unit to share your experiences in the new Science of Healthful Living Curriculum. One of the UNCG researchers will ask you to wear the accelerometer on the waist band of your shorts and on a different day, to answer questions about what you already know and what have learned, what you liked and disliked, and if you have suggestions or recommendations for changes in the new curriculum.

Did my parents say it was ok?
We will ask your parent(s) if it is ok for you to interviewed and to sign a form like this one.

Why me?
We would like you to take part because you are a student at this middle school. We selected students randomly to represent boys and girls and each grade level. In other words, we assigned each person in your class a number, put them all in a box, and drew your number out of the box.

What if I want to stop?
You can say you don't want to wear an accelerometer or be interviewed today or any day and you can drop out of the interview at any time. You also can return the accelerometer at any time during the lesson if you don't want to wear it anymore. You will not be punished if you say "no" and choosing not to participate will not affect your grade. Even if you say "yes" now and change your mind after you start wearing the accelerometer or start the interview, you can stop, and no one will be mad at you.

What will I have to do?
The UNCG researcher assigned to your class will select 6 students from your grade to be interviewed before and after each of the new *Science of Healthful Living* units. She or he will interview you individually or in a small group. You will be asked about what you already know about some of the Healthful Living topics and what you have learned in the new curriculum. The interviews will last about 30 min. and will occur during your physical education time. Your interview will be tape-recorded so we can remember what you said.

On a different day your UNCG researcher will ask you to wear a small accelerometer about the size of a pedometer on your waistband of your shorts during physical education. The accelerometer will record how hard you work that day. We will download the data into a laptop computer to be analyzed together (an average or mean) with all of the other students' data. The accelerometer measures your "metabolic rate" or how your body produces energy. To make a scientifically accurate measurement of metabolic rate, we need to measure your height and weight to enter into the computer program. We will do this privately and never reveal this information. It will only be used for the computer program to calculate your individual metabolic rate based on the accelerometer data.

Will anything bad happen to me?

No, in the interview you can say what you want. It doesn't matter if you give right or wrong answers or if you tell us you don't like the new curriculum. We want to know what you have learned and what you think. We will take the information you give us and make changes to improve the program. The accelerometer is small like a pedometer and you may even forget you have it on!

Will anything good happen to me?

You might be pleased about all the things you are learning and you might enjoy talking with your classmates about the curriculum in the small group interviews.

Do I get anything for being in this study?
No, the UNCG researcher will say, "thank you very much!" for working with us and helping us to study the new curriculum, but you will not receive any presents or prizes.

What if I have questions?
Please ask your physical education teacher or your UNCG researcher questions at any time. If you understand what we are asking you to do and want to be interviewed and wear an accelerometer, please write or sign your name below.

_____ _____

Please Write or Sign your Name on this Line Please write today's **Date** on this line.

UNCG IRB
Approved Consent Form
Valid 1/11/13 to 1/10/14

APPENDIX C

LESSON OBSERVATION CODES AND CATEGORIES

Category	Open Codes
High adherence	• Read script verbatim • Uses task cards • Teaches activity as described • Follows lesson timing • Uses journal • States the essential question • Students talking about station • Uses supplied equipment • E taught
Good delivery quality	• Teacher asks students to Think, Pair, Share • Teacher shows how to dribble • Teacher pauses after asking a question • Teacher tells student good job • Teacher checks student understanding • Teacher manages student behavior • Most students participating • Teacher answers all student questions • Teacher participates with students • Teacher enthusiastic
Low adherence	• Omitted Elaboration • Changes station to passing • Modifies task by changing equipment • Omitted two push-up types • Skip journal item • Does not refer to rubric • Physical activity homework not assigned • Students read the definitions • Omitted Evaluation • Only did one round • Whole journal completed during Engagement
Poor delivery quality	• Poor push-up form with no teacher feedback • Students talking during directions • Students appear confused • Students not following directions • Student questions not addressed • Teacher using cell phone • Teacher sitting at table checking email • Students not active