6-9-2016

A meta-analysis of the effect of the physical education learning environment on student outcomes

Stephanie Armstrong

Stephanie Lynn Armstrong
Candidate

Health, Exercise and Sport Sciences
Department

This dissertation is approved, and it is acceptable in quality and form for publication:
Approved by the Dissertation Committee:

Gloria Napper-Owen, Chairperson

Glenn Hushman

Lucretia Pence

Rock Braithwaite

A META-ANALYSIS OF THE EFFECT OF THE PHYSICAL EDUCATION LEARNING ENVIRONMENT ON STUDENT ACHEIVEMENT

by

STEPHANIE LYNN ARMSTRONG

B.S., Human Movement, Boston University, 2006
M.A.T., Elementary Education, Colorado College, 2008

DISSERTATION

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Doctor of Philosophy
Physical Education, Sport and Exercise Sciences

The University of New Mexico
Albuquerque, New Mexico

May, 2016

DEDICATION

"You are braver than you believe, stronger than you seem, and smarter than you think."

-A.A. Milne

For Mike, who believed in me even when I didn't believe in myself.

For my Mom, who is the bravest, strongest, and smartest person I know.

ACKNOWLEDGEMENTS

This dissertation represents six years of hard work by not only myself, but also an amazing group of people behind the scenes. I was lucky enough to have the absolute best support system anyone could ever ask for. Not only did the last six years represent a significant academic journey, but also a significant life journey. I was blessed to welcome two remarkable and entertaining little people into my life, Montana and Maddy. Thank you to everyone who joined me on this journey.

To my advisor and committee chair, Dr. Napper-Owen. Thank you for your patience, and insight. You were always able to see the bigger picture in my academic pursuit, and guided me to see it too. I truly can't imagine having a better mentor.

To my committee, Dr. Braithwaite, Dr. Hushman, and Dr. Pence. Your excitement and patience is what guided me through this process. I absolutely worked with an academic "Dream Team," and I cannot imagine a better group of people to work with. Dr. Braithwaite, I cannot imagine having a better mentor to help guide me through the meta-analysis process. Your countless hours on the phone and last-minute statistic sessions were invaluable. Dr. Braithwaite and Dr. Pence, thank you for your feedback along the way, and your excitement for the topic. You inspired me to constantly look at the deeper importance of the findings. Also, thank you Hong-Min Lee for your help early on in the dissertation writing process. Your hours of work on database searching and EndNote 6 troubleshooting were instrumental at the beginning of this project.

To my husband, Mike, thank you for being you. You deserve as much credit as I do for completing this milestone. I am so lucky to have you by my side, and Montana and Maddy are so lucky to have such an amazing father.

To my Mom, for being my number one supporter my entire life. You taught me to be strong, be determined, and to follow my dreams. You were and will always be my model for strength and perseverance. I am forever grateful for every sacrifice you made in your life so that Mack, Sarah and I could follow our paths.

To family and friends providing support from near and from far (the Millers, Armstrongs, Parkers, O'Briens, Sullivans, Robinsons, Teals, McGregors, Jaegers, Arandas, Abads, Karrikers, Myerses and countless others), your support means more to me than you could ever know. Your words of encouragement did not go unnoticed. I would have never completed this journey if it weren't for all of you.

A META-ANALYSIS OF THE EFFECT OF THE PHYSICAL EDUCATION LEARNING ENVIRONMENT ON STUDENT OUTCOMES

by
Stephanie Lynn Armstrong

B.S., Human Movement, Boston University, 2006
M.A.T., Colorado College, 2008
Ph.D., Physical Education, Sports and Exercise Science,
University of New Mexico, 2016

ABSTRACT

Understanding the qualities of effective teachers and finding accurate measures of student outcomes becomes paramount to knowing how to replicate student success from year to year. In 1992, The National Association for Sport and Physical Education (now SHAPE America), published the Developmentally Appropriate Practices for Children document in an effort to help practitioners better understand qualities of an effective physical education program. However, little is known about the effect that the appropriate practices in the document have on student outcomes. The best way to understand whether or not NASPE's appropriate practices have an impact on student outcomes is to investigate the body of knowledge encompassing the topic. One can begin to establish validity in the use of NASPE's appropriate practices through the use of meta-analysis research. Through a systematic review of literature, quantitative data may support the implementation of these strategies in classrooms.

The purpose of this study is to contribute to the understanding that the learning environment, as defined in the Appropriate Practices document (NASPE, 2009a, 2009b, 2009c, SHAPE 2009), has on student outcomes in physical education classes. Equally important is determining what evidence there is to support that establishing a learning environment promotes positive student outcomes in physical education.

An initial database search of terms related to the learning environment revealed 3727 citations relating to the topic. Using inclusion criteria, 19 studies (12 journal articles and 7 doctoral dissertations) were included for final analyses. A summary effect size of g=0.366 was obtained, which is considered small. Effect sizes across all studies were heterogeneous, indicating the presence of moderator variables. Moderator analyses revealed statistically significant differences for variables of safety, diversity, study design, school level, and publication type on student outcomes.

Meta-analytical results indicate a small positive relationship between the learning environment and student outcomes in the affective and psychomotor domains. More studies are needed that investigate the relationship between the learning environment and student outcomes, especially when considering outcomes in the cognitive domain. More evidence is needed to support the assumption based on professional consensus that learning environment variables affect student learning outcomes.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

Teacher effectiveness is at the forefront of educational research. Researchers are concerned with examining the degree to which teacher behaviors affect student outcomes as a measure of how effective their teaching is (Darling-Hammond, 2000). This process-product strategy can prove to be difficult in education. Some experts believe that a child's background and social context holds more weight than teacher behaviors (Coleman et al., 1966), and others argue that a large portion of student academic success can be attributed to effective teachers (Darling-Hammond, 2000). Understanding the qualities of effective teachers and finding accurate measures of student outcomes becomes paramount to knowing how to replicate student success from year to year. Replication of student success in physical education classes can be even more difficult to measure because of the lack of quantitative studies.

Finding valid measures of assessment is not the only barrier to accurately measure student outcomes in physical education. Another challenge is determining the qualities effective physical educators demonstrate that lead to student outcomes. Berliner (1976) suggests that measuring student outcomes is more complex than the process-product relationship, and that "the behavior of the student in the instructional setting" (p. 10) should also be considered. The instructional setting can be affected by factors such as class size, teacher qualifications, and school size (Darling-Hammond, 2000). This being said, the instructional setting, or learning environment, can also contribute to or hinder student academic success.

A key to measuring student outcomes in physical education relies on finding valid and reliable measures of student learning outcomes. In the general classroom today, a teacher is rated effective based on how students perform on standardized tests and a component called 'value-added.' The value-added component is defined as a measure of the estimated effectiveness of teachers or schools on student outcomes (Rubin, Stuart & Zanutto, 2004). In the physical education setting, a question remains about what is the 'value-added' component of teacher effectiveness. A historical investigation of the research literature on physical education teacher effectiveness may help to narrow the focus on the value-added component of the effective teacher of physical education.

Process-Product Research

In the 1960's and 1970's, educational researchers were concerned with the link between the individual teacher's behavior and her influence on student outcomes (Brophy & Good, 1984). Process-product research investigated the impact of the individual teacher on student mastery of the content in the curriculum. This research should not be confused with looking at teacher effectiveness, but rather at "teacher effects." (p.10). Process-product researchers were concerned with looking at teacher behaviors, and how teacher behavior affected the entire class unit. Through this line of research, it was reported that certain teacher behaviors (warmth, businesslike orientation, enthusiasm, organization, variety of materials and activities, clarity, structuring comments, probing questions, and academic activity focus) were consistently correlated with student outcomes (Rosenshine, 1971). Based on early process-product work, many assessments were designed to further understand the relationship between teacher behavior and student academic outcomes (Brophy & Good, 1984).

Theoretical Framework

Process-product supporters advocate for the relationship between teacher behaviors and student outcomes. Constructivists, however, support the idea that the social construct, or learning community, is just as important as teacher behaviors (Berliner, 1976; Bronferbrenner, 1978; Lave, 1991). Vygotsky argued that the teacher is responsible for establishing the environment in which learning is constructed (1962). Constructivists contend that learning takes place through the exchanges students have with their peers, teachers, other experts, and the environment. Teachers create the learning environment that promotes the learner's opportunities to interact with each other through discussion, collaboration, and feedback. Constructivists further defend the idea that a student's environment is important to their academic success.

Constructivism refers to the "philosophical belief that people construct their own understanding of reality" (Oxford, 1997). This knowledge centered philosophy attests that meaning is constructed based on one's interactions with the world, rather than conceding that there is simply a body of knowledge that everyone must learn. Interactions and perceptions construct realities (Warrick, n.d.). In physical education classes, students interact with the teacher, other students, and the physical environment in which class is held. Students construct meaning concerning ways of acting, interacting, and learning based on the environment established by the teacher. In light of all of the current research on teacher effectiveness, it is clear that knowing how the learning environment directly impacts student outcomes is important.

Physical Education Learning Environment

The National Association for Sport and Physical Education (NASPE) Council on Physical Education for Children (COPEC) published guidelines of appropriate and inappropriate practices for physical education teachers. The Guidelines suggest best practices that should be implemented by highly effective physical educators during instruction. In 1992, the "Appropriate Practices for Elementary School Physical Education" (NASPE, 1992) document was published, followed by the "Appropriate Practices for Middle School Physical Education" in 1995, and finally "Appropriate Practices for High School Physical Education" in 1998. It was reported that "appropriate instruction in physical education incorporates the best known practices, derived from both research and teaching experiences, into a pattern of instruction that maximizes opportunities for learning and success for all children" (NASPE, 2009a, 2009b, 2009c, p.3). Since 1992, the appropriate practices documents have been rewritten and organized into five distinct categories:

1. Learning environment
2. Instructional strategies
3. Curriculum
4. Assessment
5. Professionalism

Each of the five groups is then further divided into specific sub-categories. For the purposes of this meta-analysis project, only the learning environment category will be analyzed in depth. Under the learning environment category, seven sub-categories further illustrate the specific teacher behaviors that are considered appropriate (NASPE, 2009a, 2009b, 2009c). The sub-categories are:

1. Establishing the learning environment
2. Exercise as punishment
3. Safety
4. Diversity
5. Equity
6. Inclusion
7. Competition & Cooperation

Throughout each NASPE document, examples of appropriate and inappropriate practices found in the instructional setting are identified for each respective school level.

The Appropriate Practices Guidelines provide physical educators access to resources that offer guidance for delivering a quality physical education program; however, little is known about the evidence from which these appropriate practices are derived. Although the documents were based on expert opinion that was guided by research known at the time, there is little quantitative research in physical education based on the appropriate practices to support student outcomes. Duncan and Biddle (1974) protested that there was a tendency for educational prescriptions to be made that were not based on empirical data. Berliner (1976) echoed this sentiment in his position paper claiming that educators often commit to new practices and behaviors without evidence that they lead to student outcomes. The best way to understand whether or not NASPE’s appropriate practices have an impact on student outcomes is to investigate the body of knowledge encompassing the topic. One can begin to establish validity in the use of NASPE's appropriate practices through the use of meta-analysis research. Through a systematic review of literature, quantitative data may support the implementation of these strategies in classrooms.

Statement of the Problem

All too often educators make claims about variables that may influence student outcomes. However, much of the current physical education research utilizes qualitative findings, which may not be generalizable to the population of teachers and students in physical education. There is a need for more quantitative studies in physical education to assist teachers in making decisions that affect student outcomes.

Purpose of the Study

The purpose of this study was to contribute to the understanding that the learning environment has on student outcomes in physical education classes through the use of meta-analysis. Equally important was determining what evidence there was to support that establishing a learning environment promoted positive student outcomes in physical education. Educational theories should be grounded in measurable and replicable research.

Research Question

The purpose of this meta-analysis study was to answer the question: What are the effects of the physical education learning environment on student outcomes?

Limitations

Although the meta-analysis yielded 3727 articles, one may not conclude that this was an exhaustive search of current research articles. The search for articles was limited to five databases, which may not account for all published articles on the topic of learning environment and student outcomes in physical education. However, these databases were selected based on the prominence of physical education research being included. Included in the published research were dissertations from the ProQuest database. Dissertations are not published in peer-reviewed journals, so the data may not be available to the general public.

Although study delimitations were made in advance, there may have been keywords omitted from search criteria, and consequently the omission of articles that would have added to the validity of the findings of this project. Human error may occur when determining what articles should be kept in and which ones should be left out.

Delimitations

A team of colleagues worked together to determine delimitations for this study. We focused on research able to be generalized to a typical physical education setting using the following criteria:

1. Focus on learning environment variables for teachers and students. Exclude studies using only instructional strategies, curriculum, assessment, or professionalism variables.
2. Focus on typical physical education settings. Exclude studies conducted in laboratories, athletic teams, or facilities other than K-12 traditional physical education settings.
3. Focus on measured outcomes gain in cognitive, affective, or psychomotor domains. Study includes some form of baseline testing, or test for mediating effects, and measure growth in outcomes.

4. Focus on age group younger than 18 years old. Exclude studies involving post-secondary students, adults, or senior populations.
5. Focus on quantitative descriptive statistics and/or correlation studies. Qualitative research may accompany the quantitative research, but exclude studies that are simply qualitative in nature.
6. Focus on articles published from 1970 to present. Exclude studies conducted prior to 1970.
7. Focus on articles published in the English language. Exclude any studies published in languages other than English.
8. Focus on the databases ProQuest, PsychARTICLES, PsychINFO, SPORTDiscus, and ERIC. Exclude and studies not found in these databases.

Definition of Terms

Meta-analysis: Meta-analysis is an integrative research synthesis that attempts to "integrate empirical research for the purpose of creating generalizations (Cooper & Hedges, 1994)." The operational definition of meta-analysis in this study is "the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings (Glass, 1976)."

Learning environment: Learning environment refers to seven best practices that help students achieve and maintain physical and emotional safety while participating in physical education class. Those best practices include: competition and cooperation, diversity, establishing the learning environment, equity, exercise as punishment, inclusion, and safety. Although this list is not meant to be all-inclusive, it is the recommended list of appropriate practices for physical educators presented by NASPE (2009a, 2009b, 2009c; SHAPE, 2009).

Establishing the learning environment*:* Establishing the learning environment refers to the teacher actively planning lessons that focus on maximum learning and participation and promoting developing positive self-concept where children feel supported and able to make mistakes, and try again. Children are free from harassment from teachers and other students, and there are "fair and consistent classroom-management practices" (NASPE, 2009a, 2009b, 2009c, SHAPE 2009) in place so children follow specific behavior expectations while participating.

Exercise as punishment*:* Exercise as punishment refers to teachers promoting healthy lifestyles and encouraging exercise outside of physical education. Using exercise to punish misbehavior is not an appropriate practice (NASPE, 2009a, 2009b, 2009c; SHAPE 2009).

Safety: Safety refers to teachers using developmentally appropriate instruction, and maintaining that equipment and facilities are free of hazards. Teachers should hold up-to-date certifications in first aid, CPR, and AED. Physical education teachers should have class sizes consistent with other teachers at their school, and should monitor their classes closely (NASPE, 2009a, 2009b, 2009c; SHAPE 2009).

Diversity: Diversity refers to the teacher including all children regardless of race, ethnicity, gender, sexual orientation, physical ability, or religion. The teacher also selects activities that represent many cultures (NASPE, 2009a, 2009b, 2009c; SHAPE 2009).

Equity: Equity refers to teachers including all children equally, regardless of skill level. All students are encouraged to participate in all activities, and are not socialized into activities typically identified as “for boys” or “for girls.” Teachers also should use gender neutral language (NASPE, 2009a, 2009b, 2009c; SHAPE 2009).

Inclusion: Inclusion refers to teachers identifying students with disabilities, and following appropriate accommodation plans for their success. Teachers should also accommodate students with temporary medical limitations and for all levels of fitness (NASPE, 2009a, 2009b, 2009c; SHAPE 2009).

Competition and cooperation: Competition and cooperation refers to teachers mastery-learning environments where the focus is not always on competition. Students participate in both competitive and cooperative activities, and the emphasis is not always on winning. Students are encouraged to understand different kinds of competition, but also encouraged to set individual goals (NASPE, 2009a, 2009b, 2009c, SHAPE 2009).

Appropriate Practice: a developmentally acceptable teaching practice that is included in an “excellent” physical education program (NASPE, 2009a, 2009b, 2009c). These practices are meant to be observed by all physical education in gymnasiums, pools, and fields where physical education is taught across the United States.

Student Outcome: a measureable increase a student’s individual knowledge that escalates their preparedness for future endeavors (Student Achievement, n.d.). In physical education, knowledge can be gained in the cognitive, affective, and psychomotor domains.

CHAPTER 2

Review of Literature

Theoretical Framework

The goal of this research is to investigate the link between the physical education learning environment and student outcomes. Although this meta-analysis is quantitative in nature, central to this research is the idea that the learning environment has the capacity to impact student learning. The idea that the learning environment could affect student learning is best explained using constructivist theories.

Constructivism refers to the "philosophical belief that people construct their own understanding of reality" (Oxford, 1997). This knowledge-centered philosophy attests that meaning is constructed based on one's interactions with the world, rather than conceding that there is simply a body of knowledge that everyone must learn. It is the interactions one has with the world coupled with their unique perceptions that construct their realities (Warrick, n.d.). Constructivism is based on the philosophical branches of ontology and epistemology (von Glaserfeld, 1995).

Ontology is a branch of metaphysics that attempts to answer questions about the nature of reality, such as, "what is being?" One strand of ontology is idealism, which is grounded in the belief that there is no absolute reality (Mathis, 2011). Rather idealists, such as Plato, believe that perfect, universal ideals constitute reality (Warrick, n.d.). Idealists believe that objects are what people perceive them to be, and one person's perception varies from another person's. In contrast to idealism, realists believe that there is a true reality. Existence and independence are two cornerstones of realists' beliefs. Realists believe that objects do *exist*, and the fact that something exists is *independent* of anything anyone else thinks or says about the matter (Miller, 2014). For instance, rocks do exist regardless of what other's believe, and we construct meaning to know that a rock is indeed a rock because of its inherent rock-ness.

In contrast to ontology, epistemology is a branch of philosophy that refers to the origins, foundations, limits, and validity of knowledge itself. Epistemology attempts to answer the question, "What is knowledge?" (Warrick, n.d.). This branch of philosophy is concerned with understanding how one's knowledge and justified beliefs are created. Epistemology helps to explain how knowledge is constructed and how it is spread.

The term 'constructivist' was first credited to Giambattista Vico in 1710 in his publication entitled *De Antiquissima Italorum Sapientia* (Warrick, n.d.; von Glaserfeld, 1995). Vico was a Professor of Rhetoric at the University of Naples and taught his students that the human mind creates knowledge, and therefore people only know what their minds create (Costelloe, 2014). Vico believed that one "can rationally know only what we ourselves have made" (von Glaserfeld, 1995, p.6). He also noted that the "Latin words *verum* (the true) and *factum* (what is made) are interchangeable" (Warrick, n.d., p.9). An individual's knowledge is a combination of past and present experiences, and what is 'true' to one person varies from what is 'true' to another.

Although constructivism has roots dating back to the 1700s, Jean Piaget (1896-1980) is typically noted as the first constructivist. Piaget began his work as a biologist, and "he saw cognition as an instrument of adaptation, as a tool for fitting ourselves into the world of our experience" (von Glaserfeld, 1995, p. 14). Knowledge is not merely facts that can be taught, but also includes unique experiences constructed from the environment. Most notably, Piaget studied children and how they "built-up" their knowledge (von Glaserfeld, 1995, p.13). He maintained several assumptions about learning and development.

Piaget's first assumption was that biological and cognitive development occur through a negotiation with one's environment, and that people make their own sense about the world in which they live (Warrick, n.d.). Children have a need to organize their environment, and structures in their environment are consistently placed in higher-order systems. Humans develop more complex levels of thinking as they mature based on schemas created to help organize their information.

Piaget's second assumption refers to the need for all organisms to acclimate to their environment, which he called adaption. Adaption is composed of both assimilation and accommodation. Assimilation is the process by which people fit new information into structures that they already have (von Glaserfeld, 1995). Assimilation requires a person to put new information into schemata that already exists. The process of accommodation involves altering one's existing schemas, or creating new schemas, as a result of new information or new experiences. The works of both Vico and Piaget paved the way for various sub-categories of constructivism.

A sub-category of constructivism particularly relevant to this study is social constructivism. Social constructivism is based on the belief that knowledge is constructed through an interaction between the student, the environment, and social interactions (Warrick, n.d.). Knowledge construction is collaborative, in contrast to individualistic. Social constructivism is a reflection of Vygotsky's sociocultural theory (1978), which highlighted the role than mentors play in the construction of apprentice knowledge (Applefield, Huber, & Moallem, 2000). It is the collaboration between others, according to social constructivists, not individual cognitive investigations that ultimately leads to knowledge. In physical education classes, students interact with the teacher, peers, and the physical environment in which class is held. Students construct meaning concerning ways of acting, interacting, and learning based on the environment established by the teacher. In light of all of the current research on teacher effectiveness, it is clear that knowing how the learning environment directly impacts student outcomes is important.

Social constructivists stress the importance of the environment on student knowledge acquisition; however there is not a large body of research in the field of physical education that supports this theory. The Developmentally Appropriate Practices for Children document (NASPE, 1991) introduced physical educators to the idea that the learning environment is important. However, there is little research to support that the learning environment directly affects student learning in physical education. The purpose of this study is to determine the effect that the physical education learning environment, as defined by the Appropriate Practices documents (NASPE 2009a, 2009b, 2009c), has on student outcomes.

Developmentally Appropriate Practices

To best understand the current Appropriate Practices documents, it is important to first know their history. The following section describes the history of research on 'developmentally appropriate' practices, and how the term was adopted in the field of physical education.

Origins of Developmentally Appropriate Practices

For decades, the term 'developmentally appropriate' has been used to describe good curriculum and instructional practices for young children (Bredekamp, 1992). The term ‘developmentally appropriate’ refers to programs that are “based on knowledge of what is age-appropriate for the group of children served as well as information about what is individually appropriate” (p.31). In addition to being age-appropriate, developmentally appropriate programs take into consideration the individual needs of all children, knowing that all children develop at different rates. In the mid-1980’s the National Association for the Education of Young Children (NAEYC), a professional organization of early childhood educators, clearly defined developmentally appropriate practices for young children and adopted position statements in support of these best practices (Bredekamp, 1992). NAEYC felt the need to publish a position statement based on a growing concern regarding inappropriate practices across the country, especially in the primary grades. Knowing and understanding a child’s typical growth in all learning domains is essential for teachers while lesson planning and instructing.

The NAEYC position statement has undergone several revisions since its inception. The purpose of the position statement remains clear: “advancement in both realms: more early childhood professionals engaging in developmentally appropriate practices, and more policy makers establishing policies and committing public funds to support such practices” (NAEYC, 2009, p.23). Anyone working directly or indirectly with children should be aware of developmentally appropriate practices and should use the NAEYC document for both teaching and policy decisions.

Developmentally Appropriate Practices in Physical Education

After reviewing the NAEYC position statement, the National Association for Sport and Physical Education (NASPE) Council on Physical Education for Children (COPEC) initiated a project to create a similar document relating more specifically to physical education (NASPE, 1992). Leaders in physical education believed that inappropriate practices were not unique to classroom settings and therefore believed that NASPE should establish a position statement describing developmentally appropriate practices within physical education. With guidance from the NAEYC document, the "Developmentally Appropriate Practices for Children" (DAPE) document was published (NASPE, 1992). The intended audience for this document was "teachers, parents, school administrators, policy makers, and other individuals who are responsible for the physical education of children" (NASPE, 1992, p.5). The purpose of this document was to offer insight into practices that were both developmentally and instructionally appropriate and inappropriate for children in physical education classes.

The original DAPE document paved the way for three additional publications by NASPE, including, "Developmentally Appropriate Practice in Movement Programs for Young Children Ages 3-5" (NASPE, 1994); "Appropriate Practices for Middle School Physical Education" (NASPE, 1995); and "Appropriate Practices for High School Physical Education" (NASPE, 1998). All documents support developmentally appropriate physical education by way of knowing and understanding developmental levels and best teaching practices.

The DAPE documents are currently on their third editions and have undergone a name change from "developmentally appropriate practices" to simply "appropriate practices." This name change represents a shift in focus from child development to good teaching. The current documents focus more on desired instructional approaches than the original documents. All current editions of the document include a variation of the statement, "appropriate practices in physical education incorporates the best-known practices, derived from both research and teaching experiences, into a pattern of instruction that maximizes opportunities for learning and success of all children" (NASPE, 2009a, 2009b, 2009c). This statement acknowledges that the Appropriate Practice (AP) documents are not explicitly research based, but based on conventional wisdom and expert opinions that were guided by research. Duncan and Biddle (1974) protest that there is a tendency for educational prescriptions to be made that are not based on empirical data. Berliner (1976) echoes this sentiment in his position paper claiming that educators often commit to new practices and behaviors without evidence that they lead to student outcomes. Although NASPE used the disclaimer that the AP documents were not exhaustive, they do represent an extensive list of best practices in physical education. If practitioners are using these documents to guide their instruction, then there should be evidence (in the form of cited primary sources) that the conventional wisdom used to write these documents leads to student outcomes.

In 2009, the American Alliance of Health, Physical Education, Recreation and Dance (now SHAPE America) published a side-by-side comparison of all sections and sub-sections (Appendix A) as an addition to the traditional elementary, middle, and high school documents. This comparison includes descriptions of appropriate practices for each level of education. The language is essentially the same across the document, with slight differences based on the developmental level of the particular group of students. For instance, under the Curriculum section, sub-category 3.1.3: elementary and middle school teachers should ensure, "each lesson is designed to meet program goals as stated in a published scope and sequence," while high school teachers should ensure, "instruction follows a scope and sequence that is designed to scaffold prior learning and development mature forms of skills and strategies" (SHAPE, 2009). Original documents listed appropriate and inappropriate practices in one large list, while recent editions contain five common categories along with subcategories (Figure 1). For the purposes of this project, only the Learning Environment category, along with its sub-categories, will be studied.

Figure 2.1. *Overview of appropriate practices included in 3rd edition documents*

The **Learning Environment** section includes:
- Establishing the Learning Environment
- Exercise as Punishment
- Safety
- Diversity
- Equity
- Inclusion
- Competition & Cooperation

The **Instructional Strategies** section includes:
- Expectations for Student Learning
- Class Organization
- Class Design
- Learning Time
- Maximum Participation
- Teaching/Learning Styles
- Teacher Enthusiasm
- Success Rate
- Teacher Feedback
- Technology Use

The **Curriculum** section includes:
- Productive Motor Skill Learning Experiences
- Concept Knowledge
- Regular Participation
- Developing Health-Related Fitness
- Self-Responsibility & Social Skills
- Valuing Physical Activity
- Interdisciplinary Instruction
- Special Events

The **Assessment** section includes:
- Assessment Use
- Variety of Assessments
- Fitness Testing
- Testing Procedures
- Reporting Student Progress
- Grading
- Program Assessment

The **Professionalism** section includes:
- Professional Growth
- Professional Learning Community
- Advocacy

Learning Environment

The learning environment category was selected for this meta-analysis because it was the first category in the AP document. Essentially a meta-analysis study could be completed on each of the five separate categories in the document. The term 'learning environment' means something different in different fields of education and also has different meanings within fields. The following section sheds light on the history of learning environment research, as well as introduces the definition of learning environment that is used to guide this research project.

History of Learning Environment Research

To understand the importance of the learning environment on student outcomes, one should first understand school-effects research. School-effects research is concerned with factors affecting academic outcomes, such as the school climate (Miskel & Ogawa, 1988). School climate is often defined as the norms shared by students, teachers, or administrators at a school (Niebuhr, 1995). The purpose of school-effects research is to determine what effects a school has on student outcomes versus the effects of a child's family (Rutter, 1983). Researchers attempt to answer questions of whether schools "matter" and whether schools can affect student's behaviors and outcomes. James Coleman et al. (1966) indicated that schools made little difference in student outcomes. This sentiment was echoed by other researchers who believed that schooling could not compensate for inequalities in society and family inequality (Bernstein, 1970; Bowles, 1972; Brophy & Good, 1984). Copperman (1978) wrote that "the American educational system perpetrates a hoax on its students and on their parents" (as cited in Good, Biddle & Brophy, 1983, p.3), alluding that students do not learn anything from schools. This very radical belief was contrasted by supporters of the educational system.

Supporters of the educational system attest that school-effects do contribute to student outcomes and that even economically disadvantaged students can be successful in schools (Rutter, Maughan, Mortimore, Ouston, & Smith, 1979). Brookover and Lezotte (1979) found that school climate was significantly associated with academic outcomes, regardless of student inequalities. More recently, Niebuhr (1995) found that the school climate plays a larger role in academic outcomes than individual student motivation.

School-effects researchers neglected to look at the contributions of individual teachers in the schools (Brophy & Good, 1984). Not only do schools play a role in student outcomes, so too do individual teachers. Teacher-effect research (or process-product research) is concerned with looking at teacher behaviors and how teacher behaviors affect the entire class unit. Notably, Rosenshine (1971) reported that there were some teacher behaviors that consistently led to gains in academic outcomes. Positive correlations were found between "teacher warmth, businesslike orientation, enthusiasm, organization, variety in materials and academic activities, high frequencies of clarity, structuring comments, probing questions asked as follow up to initial questions, and focus on academic activities" (Brophy & Good, 1994, p.9). The context in which these behaviors were used was equally important. For instance, counting the number of teacher probing questions did not yield the same results as looking at comparable contexts when probing questions were used (Brophy & Good, 1984). It was often difficult to interpret beneficial teacher behaviors because of the fluid nature of classrooms. Based on early process-product work, many assessments were designed with the intent to measure teacher effectiveness in relation to student outcomes (Brophy & Good, 1984).

A notable framework currently used for evaluating teacher effectiveness is Charlotte Danielson's Framework for Teaching (2007), as provided by the Danielson Group. The mission of the Danielson Group is "to advance the understanding and application of Charlotte Danielson's concepts in the educational community, connect them to other areas of knowledge, and enhance the professional practices of educators to positively impact student learning" (https://www.danielsongroup.org/about/). Danielson's Framework for Teaching is a research-based tool that is divided into 22 components, which are clustered into four domains, including: planning and preparation, classroom environment, instruction, and professional responsibilities. The document is "based off of the Praxis III criteria developed by the Educational Testing Service (ETS) after extensive surveys of research literature, consultation with expert practitioners and researchers, wide-ranging job analyses, summaries of the demands of state licensing programs, and fieldwork" (Danielson, 2007, p.183). Danielson initially wrote her framework because she saw a need for educators, both novice and experiences, to have a "road map through the territory, structured around a shared understanding of teaching" (p.2). Her intent was to help provide professional educators with a common language based on research of best practices. Early process-product research and work on identifying effective teaching practices (Gage, 1977; Wittrock, 1986) influenced the decision to create the Framework for Teaching document.

For the purpose of this study, Domain 2: The Classroom Environment, is especially important because it represents a research-based synthesis of what the learning environment looks like in general education. Danielson (2007) attests that "attention to routines and procedures, the physical environment, and the establishment of norms and expectations for student behavior are prerequisites for good instruction" (p. 187). In addition, the classroom learning environment should include effective classroom management strategies, as well as an atmosphere of respect, caring, and commitment to work. Research that supports the need for the learning environment to be included in this teacher effectiveness framework includes the works of Evertson and Harris (1992), Jensen (1998), Tomlinson (1999), and Whitaker (2004). Evertson and Harris (1992) and Jensen (1998) supported the importance of the learning environment by documenting the need for teachers to establish routines and procedures in order to reduce stress on students. Tomlinson (1999) and Whitaker (2004) supported the importance of the learning environment by highlighting how the individual needs of all students should be respected. In all, Domain 2 includes five components pertaining to the classroom environment (Figure 2). These five components include: (a) creating an environment of respect and rapport, (b) establishing a culture for learning, (c) managing classroom procedures, (d) managing student behavior, and (e) organizing physical space. The Framework for Teaching can be used by any educational practitioner, regardless of content area.

Figure 2.2: *Domain 2: components, and elements from the Framework for Teaching document*

Domain 2: The Classroom Environment

Component 2a: Creating an Environment of Respect and Rapport

- Teacher interaction with students
- Student interactions with other students

Component 2b: Establishing a culture foe Learning

- Importance of content
- Expectations for learning and achievement
- Student pride in work

Component 2c: Managing Classroom Procedures

- Management of instructional groups
- Management of transitions
- Management of materials and supplies
- Performance of noninstructional duties
- Supervision of volunteers and paraprofessionals

Component 2d: Management of student Behavior

- Expectations
- Monitoring student behavior
- Response to student misbehavior

Component 2e: Organizing Physical Space

- Safety and accessibility
- Arrangement of furniture and use of physical resources

As research progressed from school-effects to teacher effects (process-product), then to teacher effectiveness, it is clear that the learning environment has been an important consideration in the field of education for many decades. Understanding how exactly to establish an effective learning environment and how the learning environment impacts student outcomes is important. Although other content areas have tried to establish a connection between the learning environment and student outcomes (Brookover and Lezotte; 1979; Niebuhr, 1995), the link in physical education is less clear.

Physical Education Learning Environment

Danielson's Framework for Teaching (2007) can be used to guide instruction in all content areas, however, the AP documents (NASPE, 2009a, 2009b, 2009c; SHAPE 2009) specifically outline practices in the physical education discipline. Under the learning environment category, seven sub-categories further illustrate the specific teacher behaviors that are considered appropriate in physical education (NASPE 2009a, 2009b, 2009c; SHAPE, 2009). The sub-categories are:

1. Establishing the learning environment
2. Exercise as punishment
3. Safety
4. Diversity
5. Equity
6. Inclusion
7. Competition & Cooperation

Since definitions of the learning environment vary between (and even within) disciplines, the language from the Appropriate Practices documents will be used to define the physical education learning environment. Next, all seven learning environment sub-categories are defined using language from the AP document itself.

Establishing the learning environment. When physical educators establish the learning environment, they actively plan lessons that focus on maximum learning and participation for all children. Children experience a positive learning environment where they feel supported and able to make mistakes. Students do not experience harassment from teachers or other students, and are provided consistent classroom-management practices (NASPE 2009a, 2009b, 2009c; SHAPE, 2009), where specific behavior expectations are established and consequences are in place. Students develop positive self-concept, are encouraged to focus on intrinsic incentives (NASPE 2009a, 2009b, 2009c; SHAPE, 2009).

Exercise as punishment. Exercise as punishment refers to teachers promoting healthy lifestyles and encouraging exercise outside of physical education (NASPE 2009a, 2009b, 2009c; SHAPE, 2009). Teachers do not use exercise as a consequence.

Safety. According to the AP document, the learning environment is considered safe when teachers provide developmentally appropriate instruction, maintain equipment in excellent condition, and keep facilities free from hazards. Teachers ensure student safety by staying up-to-date on their certifications in first aid, CPR, and AED; by having class sizes consistent with other teachers at their school; by specifically educating students about safety; by practicing emergency action plans; and by monitoring their classes closely (NASPE 2009a, 2009b, 2009c; SHAPE, 2009).

Diversity. The positive learning environment is established when the physical education teacher plans and delivers instruction that includes all children, regardless of race, ethnicity, gender, sexual orientation, physical ability, or religion. The teacher selects activities that represent the interests of the varied cultures of the school community. Differences are acknowledged and appreciated (NASPE 2009a, 2009b, 2009c; SHAPE, 2009).

Equity. Equity in the learning environment refers to the equal inclusion all children regardless of skill level. An equitable learning environment is one where students are encouraged to participate in all activities and are not socialized into activities stereotypically identified as “for boys” or “for girls.” Teachers also promote equity by using gender neutral language (NASPE 2009a, 2009b, 2009c; SHAPE, 2009).

Inclusion. In an inclusive learning environment the teacher identifies students with disabilities and follows appropriate accommodation plans for their success. The teacher also accommodates students with temporary medical limitations and differing levels of fitness and skill abilities (NASPE 2009a, 2009b, 2009c; SHAPE, 2009).

Competition and cooperation. Competition and cooperation refers to the provision of mastery-learning environments where the focus is not always on competition. Students are involved in both competitive and cooperative activities, where the emphasis is not always on winning. Students are encouraged to understand different kinds of competition, and also encouraged to set individual goals (NASPE 2009a, 2009b, 2009c; SHAPE, 2009).

Student Outcomes

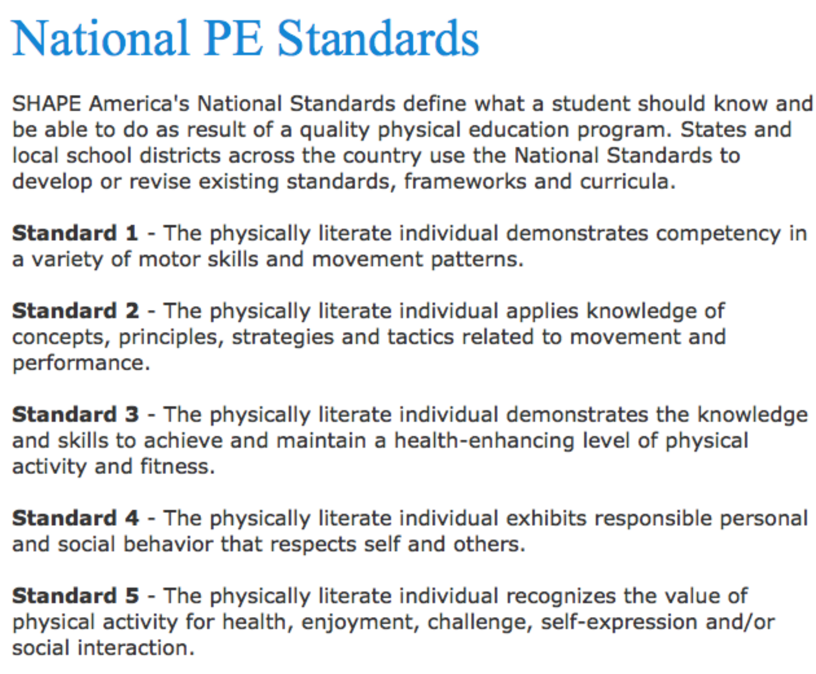
Central to this research is the idea that the learning environment has the capacity to impact student learning. In physical education environments, there are various ways to measure student growth. The final section in the review of literature presents an introduction to assessment in physical education. Also, various assessments that can be used to assess student outcomes in the psychomotor, cognitive, and affective domains are addressed.

Measuring Student Outcomes in Physical Education

The Society of Health and Physical Educators (SHAPE, 2015) and the National Board for Professional Teaching Standards (NBPTS, 2001) call for regular, quality student assessment to guide instruction, to align programs with standards, and to document student growth. One goal of quality assessment is to ensure that students are held accountable for learning based on the course outcomes (Lund, 1992; Melograno, 2007; Mercier & Doolittle, 2013). Another goal for assessment is to ensure that teachers are accountable for monitoring student growth (Mercier & Doolittle, 2013). There are currently no standardized assessments in physical education that are mandated to be used nationally, so physical educators are often free to choose whichever assessments they prefer. Assessments should be selected to reflect growth in all of the learning domains. Assessment is in integral part of an effective physical education program (Lund, 1992).

Assessing student outcomes in physical education is the responsibility of individual teachers, schools, and school districts (Mercier & Doolittle, 2013). National and state physical education standards should provide scaffolding for student-learning objectives at the individual school level. Teachers should hold their students responsible for learning in the cognitive, affective, and psychomotor domains to ensure their instruction is aligned with the national standards (Figure 3). Assessment in physical education class can take many forms, including formative and summative assessments. Formative assessments should be used at the lesson level to check how well students have learned lesson objectives (Melograno, 2007). Summative assessments, on the other hand, focus on student growth over unit, semester, or yearly goals (Melograno, 2007; Mercier & Doolittle, 2013). Information from formative assessments should be used to provide student feedback, while information from summative assessments should be used to determine student grades (Melograno, 2007).

Figure 2.3: *SHAPE America's National Standards for physical education*



National PE Standards

SHAPE America's National Standards define what a student should know and be able to do as result of a quality physical education program. States and local school districts across the country use the National Standards to develop or revise existing standards, frameworks and curricula.

Standard 1 - The physically literate individual demonstrates competency in a variety of motor skills and movement patterns.

Standard 2 - The physically literate individual applies knowledge of concepts, principles, strategies and tactics related to movement and performance.

Standard 3 - The physically literate individual demonstrates the knowledge and skills to achieve and maintain a health-enhancing level of physical activity and fitness.

Standard 4 - The physically literate individual exhibits responsible personal and social behavior that respects self and others.

Standard 5 - The physically literate individual recognizes the value of physical activity for health, enjoyment, challenge, self-expression and/or social interaction.

SHAPE (2013) emphasizes that all students should be educated in the knowledge, skills, and confidence required for a health-enhancing, physically active lifestyle. Both content standards and performance standards guide teachers in selecting appropriate assessments for all domains. Content standards define what students "should know and be able to do," while performance standards define "how good is good enough" in terms of student outcomes of the content standards. Content standards and performance standards support learning in the psychomotor, cognitive, and affective domains.

Psychomotor domain. Measuring growth in the psychomotor domain aligns with SHAPE America's (2013) standard one, "the physically literate individual demonstrates competency in a variety of motor skills and movement patterns," two, "the physically literate individual applies knowledge of concepts, principles, strategies and tactics related to movement and performance," and three "the physically literate individual demonstrates the knowledge and skills to achieve and maintain a health-enhancing level of physical activity and fitness."

Student outcomes in the psychomotor domain are typically related to gains in physical fitness or motor skills. Physical fitness can assessed through the use of the FITNESSGRAM (Cooper Institute, 2007) or President's Challenge (USA, 2015). Motor skills may be assessed by PE Metrics (SHAPE, 2010); South Carolina Assessment Program (SCPEAP, 2007); or The Test of Gross Motor Development 2 (TGMD-2) (Ulrich, 2000). Teachers often also create their own rubrics to assess how well students perform specific skills in "authentic" situations (Lund, 1992). Ideally, motor skill assessments should take the form of formative assessments, not summative assessments (Lund, 1992). Authentic formative assessments can be used by the teachers to provide students feedback, rather than to simply record a grade for student report cards. Not all states, counties, or cities require physical education teachers to use standardized assessments, so there is a lot of variety in tests used to measure outcomes in the psychomotor domain.

In the current study, student outcomes in the psychomotor domain were most often assessed through use of the TGMD-2 (Martin et al., 2009; Valentini 1999; Robinson et al. 2011), various sports skills assessments (Slack, 1976), and through the use of fitness technology (Hannon & Ratcliffe, 2004; Van Acker et al., 2010).

Cognitive domain. Measuring growth in the cognitive domain aligns with SHAPE America's standard two, "the physically literate individual applies knowledge of concepts, principles, strategies and tactics related to movement and performance and three, "the physically literate individual demonstrates the knowledge and skills to achieve and maintain a health-enhancing level of physical activity and fitness" (SHAPE, 2013).

To measure growth in the cognitive domain, Mercier and Doolittle (2013) recommend using PE Metrics assessments (SHAPE, 2010), and summative assessments found in Schiemer (2000); Hopple (2005); Graham, Holt-Hale, and Parker (2012); Mohnsen (2008); Darst, Pangrazi, Sariscany and Brusseau (2014); Lund and Kirk (2010); and Chepko and Arnold (2000). Personal fitness planning can also be used as an assessment for the cognitive domain. Summative assessments for personal fitness planning can be found in the New York State PE Profile (NYSED, 2007), and through the use of personal physical activity participation logs or journals (Lund & Kirk, 2010). Similar to the psychomotor assessments, many physical educators have the freedom to choose cognitive assessments they would like to use. These assessments can be teacher-created, or one of the previously mentioned assessments suggested by Mercier and Doolittle. In the current study, student outcome in the cognitive domain were not reported in any of the primary studies meeting inclusion criteria.

Affective domain. Finally, measuring growth in the affective domain aligns with SHAPE America's standard four, **"**the physically literate individual exhibits responsible personal and social behavior that respects self and others" and five, "the physically literate individual recognizes the value of physical activity for health, enjoyment, challenge, self-expression and/or social interaction" (SHAPE, 2013). Student outcomes assessed in the affective domain include knowledge of personal and social skills, as well as personal and social behavior in activity settings. Students should be able to demonstrate "positive actions and interactions… such as following directions, being fair, demonstrating proper etiquette and playership, and interacting properly with peers and instructors" (Gallo, 2003). Gallo suggests that teachers use objective measures to assess student outcomes in the affective domain. Sample affective assessments include the New York State PE Profile, standards 1A & 2 (NYSED, 2007); PE Metrics standards 5 and 6 multiple-choice tests (SHAPE, 2010); and the Affective Domain Criteria (Gallo, 2003).

In the current meta-analysis, the majority of reported student outcomes across all published research were from the affective domain. Variables included caring and other social skills (Balderson, 2006; Goudas & Magotisiou, 2009), enjoyment (Viciana, 2007; Barkoukis, 2010), and cooperation skills (Goudas & Magotsiou, 2009).

Summary

Social constructivists have worked to illustrate the importance of one's environment on their acquisition of knowledge. Students are not passive in their construction of knowledge; they make meaning out of their interactions with others, and with the environment. School-effects and teacher-effects (process-product) researchers used the foundational beliefs of constructivism to determine how exactly one's school environment or teacher plays a role in a child's academic outcomes. School-effects research found that school climate is especially important in affecting student outcomes, while teacher-effects research found that several specific teacher behaviors consistently led to student outcome gains.

The effect of the individual teacher on student development is becoming extremely important to educational policy makers. Research-based tools such as Danielson's Framework for Teaching (2007) are being used to determine the extent to which teachers are utilizing effective teaching practices. One of the four domains utilized in Danielson's framework is the learning environment. In all content areas, it is important to provide a safe and positive learning environment for students. SHAPE America's Appropriate Practices documents provide a definition for what a safe learning environment should look like in physical education class.

The purpose of this study is to contribute to the understanding that the learning environment has on student outcomes in physical education classes through the use of meta-analysis. Equally important is determining what evidence there is to support that establishing a learning environment promotes positive student outcomes in physical education.

Chapter 3

Methodology

Introduction

The purpose of this study was to contribute to the understanding that the learning environment has on student outcomes in physical education. Equally important was determining what evidence there was to support that establishing a learning environment promotes positive student outcomes in physical education. Because of the nature of the inquiry, a meta-analysis review was used to collect any quantitative data previously reported relating to the goals of this research project. This chapter describes what a meta-analysis is, along with the methods used in the present study.

What is a Meta-Analysis?

Meta-analysis is a systematic review that "refers to the statistical synthesis of results from a series of studies" (Borenstein, Hedges, Higgins & Rothstein, 2009, p.xxi), also known as an "analysis of analyses" (Glass, 1976). Gene Glass developed the technique in the 1970's. In a meta-analysis, limitations and rules are used to find and include studies in the analysis (Borenstein et al., 2009). The process of using rules and limitations helps to make the meta-analysis more objective, although objectivity cannot be achieved entirely. Interpretation of the project limitations, limits on databases searched, and inherent human error can contribute to the subjectivity of a meta-analysis. Meta-analysis relies on quantitative data that is typically reported in studies, such as mean and standard deviation. It depends on the use of effect size to measure and significant findings (Lipsey & Wilson, 2000; Borenstein et al., 2009). Varying types of statistical forms can be compared, from pre-post mean differences, to bivariate correlates

(Lipsey & Wilson, 2000). Meta-analysis can be used to synthesize evidence from interventions, support evidence-based policy, or help design new studies based on gaps in the research (Borenstein et al., 2009). Using meta-analysis allows for more objectivity than narrative research, and provides insight regarding what other bodies of literature have to say about a common topic.

Meta-Analysis in Education

The literature in the field of education provides countless strategies for teachers to consider when trying to meet the needs of all of their students. Although most schools have specific curriculums and assessments to use, teachers have many freedoms in how they deliver content to their students. There are countless anecdotes about inspirational teachers or principals; tales about great educational innovators; and claims made about the effective curriculums at magnet, charter, or focus schools (Hattie, 2009). These inspirational stories lead to an “uncoordinated acceptance of too many different innovations” (Hattie, 2009, p.2), which ultimately is not supported by evidence of student outcomes. Since there are numerous articles written about effective teaching, it is often difficult to summarize them all and determine what strategies work best, based on evidence. Meta-analyses help synthesize all of the previous research which makes it easier for educators to identify best practices, which contributes to effective teaching.

In education, there is often a reliance on “common sense” teaching strategies (Hattie, 2009). Common sense strategies are used by teachers because they make sense, such as increased academic learning time or showing respect. With the overabundance of educational research, it is understandable that teachers adopt practices that appear reasonable without investigating their link to student outcomes. Some teacher traits, such as keeping eye contact or using advanced organizers, might be correlated with increased student outcomes, but may not lead to the largest outcome gains. Best practices should be adopted if evidence shows a meaningful relationship to outcomes, not just claim to work anecdotally.

Advantages of Meta-analysis

Common sense teaching strategies are popularly used, however, they do not consider sample sizes, strength of results, or even publication bias (Borenstein et al., 2009). Lipsey and Wilson (2000) state that there are four clear advantages of the meta-analytic approach over other forms of review. These advantages include: (a) there is a clear structure, including documentation and criteria selection; (b) key findings are reported by calculating magnitude and direction; (c) the synthesized effect sizes have more statistical power than individual studies, and these effect sizes are not available through narrative summaries; (d) systemic coding and use of spreadsheets to record data is much more effective than the use of subjective coding and note-taking. Even with these clear advantages to using meta-analysis, there are also several popular criticisms to the approach.

Criticisms and Defense of Meta-analysis

There are several common criticisms to meta-analysis research, including: the "apples to oranges," "flat earth," "garage door," and "file drawer" arguments. There are, however, equally as many defenses.

Apples to oranges. A common criticism with the meta-analysis is that it compares "apples to oranges," meaning that so many different studies should not be combined to yield one result (Hattie, 2009). To this point, Glass argued that all of the articles do not need to be the same (Glass, 2000). He contended that comparing apples to apples is trivial. The variance in the articles is important. Supporters of meta-analyses argue that the comparison of different studies is the only comparison that makes sense, since similar studies do not need to be compared (Glass et al., 1981; Lipsey & Wilson, 2000). Analyzing studies that are similar is illogical since they should all lead to the same finding.

Another defense of this argument is that a meta-analysis usually involves the synthesis of research pertaining to one topic, so apples *are* being compared to apples (Light & Pillemer, 1984). Inclusion and exclusion criteria are used to ensure that the synthesis includes only articles related to a particular topic. “All studies differ and the only interesting questions to ask about them concern how they vary across the factors we conceive of as important.” (Glass, 2000, np.) One of the strengths of a meta-analysis is that although many different resources (or fruits) are brought together to answer one question, there is consistency and generalizability abstracted from those resources (Borenstein et al., 2009).

Flat earth criticism. The “flat earth” criticism was first made by Cronbach in 1982. He argued that meta-analysis research only looked at the big picture, and did not give more sophisticated answers.

> “...some of our colleagues are beginning to sound like a kind of Flat Earth Society. They tell us that the world is essentially simple: most social phenomena are adequately described by linear relations; one-parameter scaling can discover coherent variables independent of culture and

population; and inconsistences among studies of the same kind will vanish if we but amalgamate a sufficient number of studies.... The Flat Earth folk seek to bury any complex hypothesis with an empirical bulldozer." (Cronbach, 1982, p. 70.)

While the results of a meta-analysis *could* simply produce effect size, moderators are used to look at complexities and try to find meanings (Hattie, 2009). The world is full of variances, not just "crude numbers" (Glass, 2000), just as meta-analysis results are full of moderators, not simply effect size calculations.

Garbage in – garbage out. Another criticism of meta-analyses is that low quality studies are included in the synthesis (Hattie, 2009; Borenstein et al., 2009). The problem with using low quality studies is that errors in primary studies will carry over into the meta-analysis. There are two possible solutions to this criticism. The more strict approach involves setting inclusion and exclusion criteria at the onset of the study to control for study quality (Lipsey & Wilson, 2000). This strategy ensures that low quality studies are excluded from effect size calculations. A second suggestion is to include all available studies. Sometimes it is more advisable to use both high and low quality studies, "code them for the nature of the experimental design and for the quality of the study, and then use meta-analysis techniques to address whether the effects differed as a consequence of design and quality" (Hattie, 2009, p.11). This second approach stresses the importance of finding and summarizing all possible studies, regardless of the quality. When reviewing the included studies, one can determine whether the article quality should also be a moderator.

The file drawer problem. There is evidence to show that "studies finding relatively high treatment effects are more likely to be published than studies finding lower treatment effects" (Borenstein et al., 2009, p.378). Publication bias results when only articles with significant findings are being reported. Articles with adverse or low treatment effects may simply remain in a filing cabinet, never to be published. In response to this criticism, one must first admit that publication bias exists in many forms of research, not purely meta-analysis. Distinguishing between robust studies and "suspect" studies (Borenstein et al., 2009), or including unpublished articles, such as thesis and dissertations (Rosenthal, 1979), may help prevent publication bias.

Regardless of the arguments for or against meta-analysis, it a technique that is an accepted method for conducting research. Meta-analyses are used in a variety of scholarly fields. A recent search of the ERIC database revealed 2,576 documents including the term "meta-analysis," and 19 specifically relating to meta-analyses in physical education.

Effect Size Calculations

An effect size is "the degree to which the phenomenon is present in the population, or the degree to which the null hypothesis is false" (Cohen, 1988, pp. 9-10). Any effect less than or equal to .20 is considered small, less than or equal to .50 is considered medium, and less than or equal to .80 is considered large (Cohen, 1988). Effect size calculations allow the researcher to meaningfully analyze studies, even if the studies do not use the same measurement procedures (Lipsey & Wilson, 2000). In a meta-analysis, standardization of statistics is necessary, which results in interpretable results across all studies. The most common effect size statistics are based on calculations of standard deviation. Standard deviations from all representative studies are combined and operationalized. These numbers can then be compared, and means, variances and correlations can be computed (Lipsey & Wilson, 2000).

Although the standard deviation calculation is most common in meta-analysis, it is not the only way to calculate effect size. The *p*-value (used to find statistical significance), correlation statistics, and other commonly used statistics can also be used.

> "The more desirable forms (of calculating effect size) index both the magnitude and the direction of a relationship, not merely its statistical significance. In addition, they are defined so that there is relatively little confounding with other issues, such as sample size, which figures prominently in significance test results" (Lipsey & Wilson, 2000, p.5).

The research question should guide the researcher in selecting the correct process to use in the effect size calculations.

Statistical Models

The research question not only guides the researcher in selecting effect size calculation processes, but also aids in the decision of which statistical model to use. There are two primary models that may be selected when using a meta-analysis: a fixed-effects or a random-effects model.

In a fixed-effect model, the assumption is that all studies included in the analysis are reporting a common, or true, effect size (Borenstein et al., 2009). All included studies are comparing the same variables, and using the same methods. For example, all studies are reporting the use of the same drug, used on the same population, at the same dosage, for the same condition. In a fixed-effects model, the random error reported is due to sampling error or within-study variance (Borenstein et al., 2009). Study weights are used, and studies with larger samples are given a higher weight in the effect size calculation to minimize the effects of sampling error.

In contrast to a fixed-effect model, the random-effects model is used when variables across studies are not consistent. This model assumes that the effect sizes in the studies are “similar but not identical” (Borenstein et al., 2009). In this model, the same variables are not reported in each study, and the variance in the studies needs to be accounted for. For instance, the participants might be at different levels of schooling, with different learning environment variables manipulated. One must assume that the true effects are normally distributed around the mean in this model. In a random-effects model, both sampling error and between study variance are sources of error. Tests for heterogeneity should be used to determine whether the variance in the articles is due to chance. Study weights are assigned in a random-effects model to decrease the impact of the variance (Borenstein et al., 2009).

A random-effects meta-analysis was selected for this study based on the inherent variation between all gathered studies. For instance, the study participants varied in student age, gender, and school type. Also, learning environment variables and student outcomes differed from study to study. The random-effects model assumes that not all variables in the studies were controlled for (Borenstein et al., 2009). Because of these variants, the magnitude of impact of the learning environment variables differed from study to study.

Data Collection

This meta-analysis was designed to answer the question of whether the physical education learning environment affects student outcomes, and if so to what extent. A systematic approach used in conducting a regular meta-analysis was followed. After specifying the research question the following steps were followed:

1. Selected appropriate databases to search
2. Created inclusion/exclusion criteria
3. Reviewed and selected articles
4. Developed a coding form
5. Coded each included study
6. Extracted data
7. Conducted statistical analyses
8. Interpreted findings

Procedure

Database search. The search of reference databases was the exclusive method of literature collection. While database searching may not be an exhaustive method of article collection, it yields the highest results of references when trying to synthesize data (Cooper, 2010). Databases are updated frequently and include articles from public and private organizations. The use of databases also allows for the most replicable results. Although databases are constantly being updated, there can also be lag time between when an article is published, and when it appears in the database (Cooper, 2010). Nonetheless, searching databases was the best method of data collection for this specific meta-analysis. Five databases were used for the initial literature search. These databases included: ProQuest, PsychARTICLES, PsychINFO, SPORTDiscus, and Education Resource Information Center (ERIC). The ProQuest database includes mostly thesis and dissertations. Including this database in the search helped lower the bias of only using published articles in the literature review. Cooper advised not to restrict your search to published articles, since "the possibility of bias against the null hypothesis is too great" (2010, p.76).

An exhaustive search was completed in all five of the databases, *always* using the search term "physical education" along with at least one of the following search parameters: positive learning environment, management, behavior management, classroom climate, discipline, routines, grouping, physical safety, emotional safety, communication, feedback, content development, instructional strategies, learning, or outcomes. Suggestions from Cooper (2010) were followed, which were to use the AND Boolean operator, not the OR operator. This allowed for two terms to be searched simultaneously. For the initial search, a combination of phrases using both quotation marks and no quotation marks were used. For example, both: physical education and "physical education" were used as separate search terms. Often the term physical education without quotation marks would result in articles related to physical science or other topics unrelated to the purpose of this meta-analysis. A standard table was used to record all articles found in the databases according to the sets of key words used (Appendix B). This table was also used to record the number of articles that were excluded for not meeting the search term criteria. The initial database search resulted with a total of 3727 articles found. All articles were stored in a shared EndNote file using EndNote6 software.

Initially, the database search was conducted by five colleagues. At the time during which the review was conducted, two researchers were faculty members at the University of New Mexico, one was a faculty member at Humboldt State University, and two were PhD students at the University of New Mexico (including the dissertation author). The faculty member from Humboldt had the most expertise of the group and guided many of the decisions in the literature retrieval and coding processes.

Inclusion criteria. As with all forms of systematic reviews, a set of inclusion and exclusion criteria, also known as conceptual definitions, was specified (Cooper, 2010). Inclusion criteria direct the types of studies that will be included in the review. Inclusion at the early stages of the meta-analysis should remain open to interpretation and should err on being overly inclusive (Cooper, 2010). Being overly inclusive at early stages of analysis will avoid leaving out important articles. Conceptual definitions should be revisited later in the analysis, and more precise definitions can be made.

For the purpose of this meta-analysis, criteria for the initial screening included:

1. Study was conducted in a Physical Education context with students ranging from 5 to 18 years of age.
2. Study reported outcomes for teachers and/or students
3. Study included learning environment variables as defined by NASPE's appropriate guidelines for creating a learning environment.
4. Study included quantitative descriptive and/or inferential statistics to be able to provide an estimate of effect size.

5. Articles from 1970 published in English Language Journals.
6. Articles were published in English.

Reviewing and selecting articles for inclusion. There are several strategies one can use to determine the inclusion or exclusion of an article. Simply reading the title of an article can help provide insight to whether or not it should be included, but can also lead an article to be excluded prematurely. Reading an entire article provides greater insight than the title, but is time consuming. Cooper (2010) suggests reading both the title and abstract as a more reliable source of determining inclusion or exclusion.

The technique used in this analysis was a title and abstract review. After review of the title and abstract, a decision was made regarding whether or not to retrieve the full document. This approach enabled the reviewers to have a clearer understanding of the study. Also, since the ProQuest database was used which primarily has thesis and dissertations, titles and abstracts were more readily available than full text copies.

After the initial database search was completed, the two PhD students separately conducted a brief title and abstract review of the 3727 articles. Because of the large number of articles found, it was necessary to complete a concise review to eliminate any articles that obviously did not meet the needs of the study. Having two students complete the review increased the inter-rater reliability. All 3727 articles were moved into one of the four following EndNote6 folders:

1. In
2. Out: No student outcomes
3. Out: Not physical education
4. Out: Insufficient data

On some occasions an abstract did not provide all necessary information or an abstract was not readily available. Therefore the decision was made to be more inclusive then exclusive at this stage to avoid missing significant documents. After combining both students' findings and excluding duplicates, a total of 669 articles remained to be further investigated.

EndNote6 has a feature that enables the user to check for duplicate articles and delete duplicate publications. This duplicate feature works well if the article appears with the *exact* same title with variance in punctuation or capitalization. Because of this, it was decided to manually review the list of 669 articles to check for duplicates that the software missed. After manually checking for duplicates, the final list of 564 articles was used for the next phase of analysis.

In/out form. The physical education learning environment In/Out Form (Appendix C) was used during the next stage in the review to determine if articles should be included in the final analysis. During this step, all 564 article abstracts were read thoroughly to determine if the article clearly met the limitations of the study. If there was not enough information in the abstract, then the methods and results sections of the articles were also read. If a full-text version was not readily available, then articles were assigned to a "maybe" folder for further review. Limitations at this stage remained the same as the previous title and abstract review.

At this point in the literature review, the list of 564 articles was divided evenly among all 5 colleagues. All colleagues were at various stages of completion of this stage when the dissertation author took over the project. Upon reviewing the available abstracts of the 564 articles, 75 were labeled for full retrieval.

Gathering full articles. Following the title and abstract review, several steps were taken to obtain PDF copies of the articles. All articles found in databases other than ProQuest were found by searching the University of New Mexico library system or through Google searches. If items were not available through the library, then an inter-library loan was requested. Most articles arrived via PDF, while others arrived as microfilm or microfiche and needed to be scanned into PDFs.

Most articles from the ProQuest database were requested through inter-library loan (ILL) at the University of New Mexico's Zimmerman Library, while others were downloaded as full-text through the library website. One dissertation was sent through mail (through some detective work by the dissertation chair) when it couldn't be found through ILL. Another dissertation could not be acquired through ILL or communication with the author, so it was ultimately left out of the meta-analysis, leaving 74 total articles for further review. The abstract, methods, and results sections of the dissertations were all scanned by hand from bound copies of the dissertations to make PDF's for analysis.

Data Analysis

The following section describes the analysis employed after gathering data from the 75 remaining. Analyses of the data included calculation of kappa coefficient, calculations of effect size, tests of heterogeneity, moderator variables, and publication bias.

Coding articles and kappa coefficient. A data coding form was developed by the researcher (Appendix D) to code articles for the meta-analysis. The design for the extraction form was based on (1) methodology used, (2) participant information, (3) study type. Methodological features provided details concerning methods used during the study and included: (a) Research design (experimental, descriptive, longitudinal, other); (b) Outcomes (psychomotor, cognitive, affective, multiple outcomes; (c) Design (between subject, within subject, other); (d) Time frame (0-3 months, 4-6 months, greater than 6 months); (e) Manipulation check (yes, no); (f) Exercise as punishment (children encouraged to participate in PA outside of school, uses exercise as punishment, not reported); (g) Establishing the learning environment (maximum participation, respect and support atmosphere, developing positive self-concept, free of criticism and harassment, emphasis on intrinsic incentives, fair and consistent management, inappropriate behavior consequences, not reported); (h)Safety (teacher teaches safety, practicing emergency action plans, activity for developmental level, teacher CPR, first aid, AED qualifications, facilities and equipment maintained and inspected, class size consistent with other subject areas, teacher monitors class, not reported); (i) Diversity (inclusive and supportive environment, differences acknowledged and appreciated, culturally diverse environment, not reported); (j) Equity (all children have equal opportunities to participate, boys and girls are encouraged and supported toward outcomes, teacher uses gender-neutral language, not reported); (k) Cooperation and competition (Different kinds of competition are taught, mastery-learning environment, acceptance of cooperative and competitive student preferences, not reported); and (l) Inclusion (Teacher uses IEP or school accommodations, lessons adapted for overweight children, adaptations for temporary medical limitations, not reported). Participant features included (m) School level (elementary school, middle school, high school); (n) Sex (male, female, combined); (o) Geographic location; and (p) Type of school (public, private, combination, other). Study type included (q) Reporting method (self-report, objective report, combined reporting); and (r) Publication type (journal article, dissertation or thesis, other).

A random sample of 35 articles were re-coded following a 2-month period, which provided the researcher with a kappa coefficient. This provided the researcher with a measure of intra-rater reliability. Cohen's kappa is used to measure reliability, and adjusts for the chance rate of agreement. Kappa is defined as "the improvement over chance reached by the coders" (Cooper, 2010, p.103). Typically, a kappa of < 0 represents less than a chance agreement, 0.01-0.20 represents slight agreement, 0.21-0.40 represents fair agreement, 0.41-0.60 represents moderate agreement, 0.61-0.80 represents substantial agreement, and 0.81-0.99 represents an almost perfect agreement (Viera & Garrett, 2005).

Effect size. Data for effect size calculations were entered into Comprehensive Meta-Analysis (CMA) version-2 software (Borenstein, Hedges, Higgins & Rothstein, 2005). The statistics gathered to calculate effect sizes included (1) means and standard deviations; and (2) between group comparisons based on means and standard deviations, t values, F values, and *p* values. Correlation coefficients were ultimately not included since they did not provide information about the effects of a learning environment intervention. If multiple outcomes were reported in a single study, then the average score was used, resulting in one overall combined calculation (Borenstein et al., 2009). Effect sizes were calculated using Hedge's g.

To guarantee data extraction quality, a dissertation committee member worked closely with the researcher, and provided number checks. A sample of studies was selected to be coded twice to insure intra-rater reliability. When disagreements on items were found, the researcher consulted committee members to arrive at a consensus. Cohen's (1988) criteria were used, with .2 as small, .5 as medium, and .8 as large effect size. These categories are relative, based on the methods that are utilized within studies. However, the criteria provide a sense of the magnitude of the effect, which is interpreted based on the data from which it is calculated.

Heterogeneity. Next, heterogeneity tests were conducted to determine the variation in the effect sizes (Borenstein et al., 2009). Variance always exists in studies, even when measuring the same variable with the same population. Heterogeneity tests measure the differences between studies that are *not* related to chance. First, the Q statistic was computed, followed by Tau squared, Tau, and finally I-squared.

The Q statistic is utilized for significance testing across a number of effect sizes and is "a ratio of between-studies variance to within-studies variance" (Rosenthal, Hoyt, Ferrin, Miller, & Cohen, 2006, p. 243). Ultimately, the Q statistic is used to determine whether or not there is heterogeneity, and not to quantify the amount of heterogeneity. The process of calculating the Q statistic involves computing the deviation of each effect size from the mean, squaring it, and weighting it by the inverse-variance for the study (Borenstein et al., 2009). These values from all of the studies provide the weighted sum of squares (*WSS*), or the Q statistic. A significant Q statistic (based on the corresponding p-value) represents heterogeneity in the studies.

After calculating the Q statistic, Tau squared (T^2) and Tau (T) were calculated. Both of these measures use the same scale as effect size in their calculations. T^2 represents the variance of the true effect sizes (Borenstein et al., 2009), while T is the estimate of the standard deviation of the true effect size. These statistics are used to make a value judgement on the range of the effect size.

Unlike T^2 and T, the I^2 statistic is not based on the same scale as effect size. The I^2 statistic represents "the ratio of true heterogeneity to total variance across the observed effect estimates" (Borenstein et al., 2009, p.117). In other words, it is the percentage of variance across studies. I^2 is computed on a range from 0-100%, and is not directly affected by the number of included studies (Borenstein et al., 2009). If I^2 is near zero, then almost all of the observed variance in the studies is minimal. Low, medium, and high benchmarks for I^2 are 25%, 50%, and 75% respectively (Higgins, Thompson, Deeks & Altman, 2003).

Following heterogeneity tests, moderator analyses were used to determine if the learning environment variable, learning outcome, student grade level, or student gender resulted in different effect sizes (Borenstein et al., 2009).

Moderator variables. In a random-effects meta-analysis, multiple moderators of effect size can be present across studies. To determine the overall effect of each individual moderator (such as student grade level or learning domain), one can control for the moderator and determine its effect on the total effect size (Cooper, 2010). When heterogeneity tests determine that there are moderate or high levels of between-study variance, then moderator analyses should be conducted.

In the current study, moderator analyses were conducted for all variables recorded in the final coding form (Appendix D). The impact of the learning environment on the psychomotor, cognitive, and affective domains were controlled for, and their impact on the total effect size was calculated. Future goals for this research include looking at the impact of all possible moderator variables on the overall effect size.

Publication bias. Finally, publication bias was addressed. Published studies are more likely than unpublished studies to report high effect sizes (Borenstein et al., 2009). Similarly, published studies are more likely than unpublished studies to be reflected in a meta-analysis. The tendency for published studies with larger effect sizes to be included in meta-analysis is known as publication bias.

Several strategies were used to counteract bias in this study. First, the database ProQuest was used, which includes unpublished dissertations and theses. Inclusion of both published journal articles and unpublished doctoral dissertations enhanced literature quality and reduced publication bias. Quality judgements on the studies themselves were not made. Cooper (2010) suggests that it is best to include all studies according to the pre-set inclusion criteria. The inclusion of both high and low quality studies helps reduce publication bias, since all studies have equal opportunities to be included.

Next, publication bias was evaluated by constructing a funnel plot, calculating Fail Safe *N*, and utilization of Trim and Fill (Borenstein et al., 2009). The decision to use all three methods was based on the desire to find consistency.

A funnel plot displayed the relationship between study size and effect size (Borenstein et al., 2009). The use of a funnel plot provides an initial observation of the distribution of studies around the mean. In this method, effect size was plotted on the X axis, while standard error was plotted on the Y axis. In the absence of publication bias, the studies would be scattered symmetrically around the mean effect size (Borenstein et al., 2009).

Fail Safe *N* was used to compute how many missing studies were needed to reject the null hypothesis of no effect (Borenstein et al., 2009). To calculate Fail Safe *N*, an assumption was made that the effect size of all missing studies was zero.

The third evaluation of bias was made using Duval and Tweedie's Trim and Fill. This process provided the best estimate of an unbiased effect size by removing extreme small studies from the positive side of a funnel plot, and re-computing until the funnel plot was symmetric (Borenstein et al., 2009). This process of trimming and filling resulted in a new, unbiased estimate of effect size.

Summary

This chapter described the characteristic of a meta-analysis, and explained the steps performed in the current project. Using five databases (including both published and unpublished primary studies), setting inclusion criteria, and utilizing both inter-rater and intra-rater reliability helped to prevent publication bias and increase the validity of the study. The next chapter presents the results from this study.

Chapter 4

Results

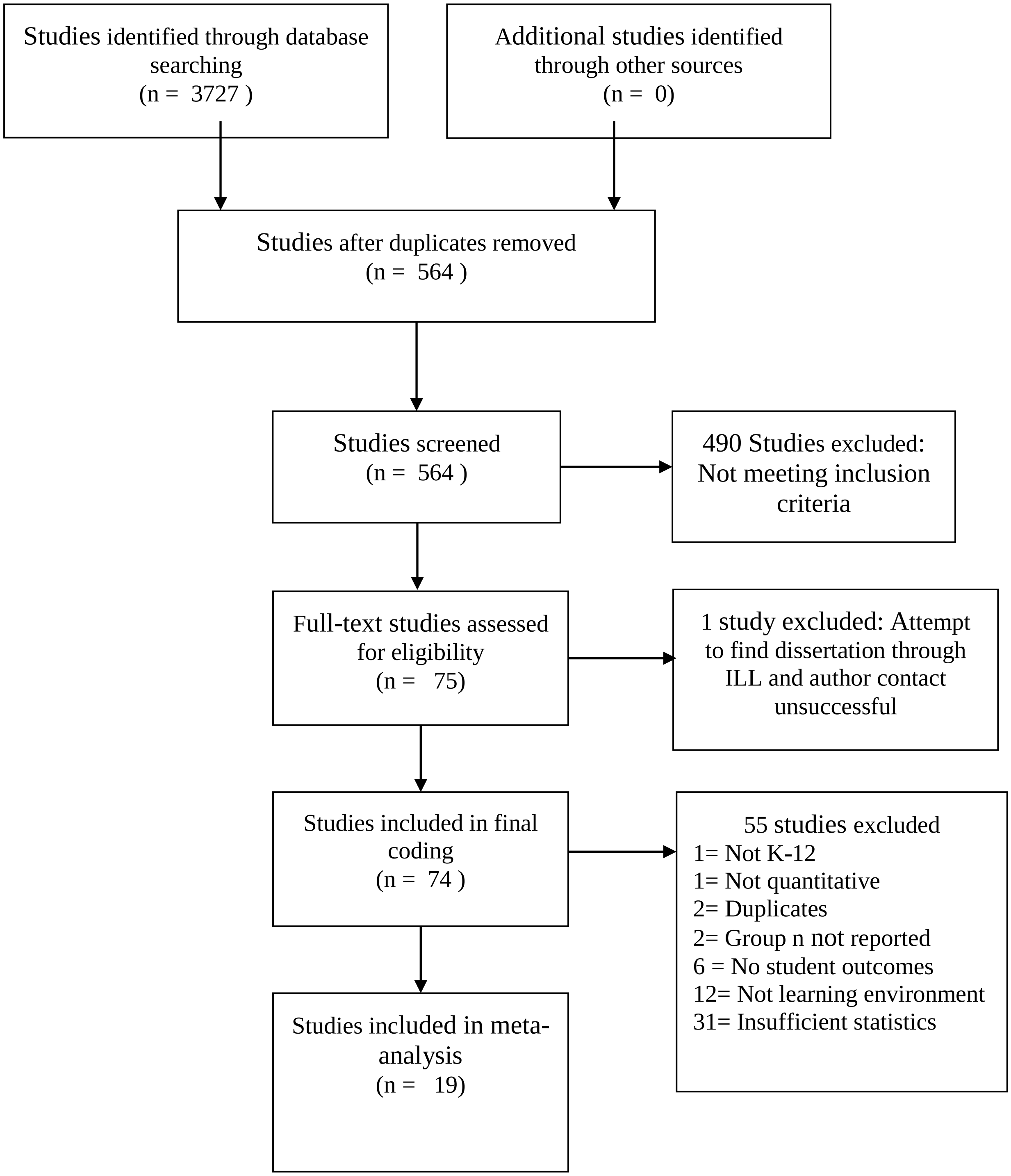
The purpose of this study is to determine the overall relationship between the physical education learning environment and student outcomes in the cognitive, affective, or psychomotor domains. There were a total of 19 studies that included 2294 participants meeting inclusion criteria. This chapter presents the results of the systematic analysis that responds to the research question. An overview of the included studies will be presented, followed by the kappa calculation, effect size calculations, outlier analysis and publication bias, tests for heterogeneity, results of the moderator analyses, and finally results from the outcome analysis.

Overview of Included Studies

Detailed descriptions of the methods used in the literature search were described in chapter 3. Figure 4.1 displays a flowchart that visually represents the process and outcomes of the literature search.

An initial database search produced 3727 total articles that met initial search term criteria. Following a brief title and abstract review, 564 that met all inclusion criteria were kept. Following a full abstract review, 74 articles remained for coding. After coding and data extraction 55 articles were excluded due to being duplicates (both dissertation and journal article were included), not reporting student outcomes, not using a K-12 population, not including quantitative statistics, not reporting learning environment variables, or not including statistic needed to calculate effect size. After the full text review and document coding, 19 articles (7 dissertations and 12 journal articles) met all inclusion criteria and provided adequate statistics for effect size calculations.

Figure 4.1: *Flowchart of the data inclusion process*



Measures of Student Outcomes

Thirty-six different instruments and researcher-made assessments were used to measure student outcomes in the 19 primary studies included in the analysis (Table 4.1). The Learning and Performance Orientations in Physical Education Classes Questionnaire (LAPOPECQ; Papaioannou, 1994) was the most widely used assessment to measure affective domain outcomes. The Test of Gross Motor Development (Ulrich, 2000) was the most widely used assessment to measure psychomotor domain outcomes. Three assessments were created by the researcher, while the other 32 assessments were previously developed assessments.

Table 4.1: *Assessments used to measure student outcomes*

	Name of Instrument	First Author of Primary Study
Affective Domain	Sociometric Rating Scale Technique (Frederickson & Furnham, 1998)	Andre et al., 2011
	Teaching personal and Social Responsibility Scale (Hellison, 2003)	Balderson, 2006
	Task and Ego Orientation in Sport Questionnaire (TEOSQ; Duda and Nicholls, 1992)	Barkoukis et al., 2010 Barkoukis, 2008
	Intrinsic Motivation Inventory (Ryan, 1982; McAuley, Duncan, & Tammen, 1989)	Barkoukis, 2008
	Learning and Performance Orientations in Physical Education Classes Questionnaire (LAPOPECQ; Papaioannou, 1994; Spanish version Jimenez, 2004)	Barkoukis, 2008 Barkoukis et al., 2010 Viciana et al., 2007 (Used Spanish version)
	Physical Education Trait Anxiety Scale (PETAS; Barkoukis, 2001)	Barkoukis, 2008
	Inventory developed by Duda, Fox, Biddle, and Armstrong (1992) to measure children's affective responses in sport	Barkoukis et al., 2010
	Perception of Success Questionnaire POSQ (Roberts & Balague, 1991 Spanish version Cervello, Escarti,& Balague, 1999; Cervello & Santos-Rosa, 2000)	Cramer, 2000 Viciana et al., 2007 (Used Spanish Version)
	Carlson's Physical Education Survey (Carlson, 1995)	Cramer, 2000
	Multidimensional Scaled of Perceived Self-Efficacy (MSPSE; Bandura, 1990, 2001)	Escarti et al., 2010A

	Name of Instrument	First Author of Primary Study
Affective Domain Continued	Multisource Assessment of Children's Social Competence (MASCS; Junttila, et al., 2006; Greek version Magotsiou, Goudas, & Hasandra, 2006)	Goudas & Magotsiou, 2009
	Feelings Toward Group Work scales (Cantwell & Andrews, 2002; Greek version Goudas, Magotsiou, & Hatzigeorgiadis, 2009)	Goudas & Magotsiou, 2009
	Likert scale of student's views toward team sports and coed versus single gender physical education settings	Hannon & Ratcliffe, 2004
	Severy' s (1975) questionnaire instrument	Polvi & Telama, 2000
	An instrument constructed based on theories of social behaviors	Polvi & Telama, 2000
	Classroom Life Questionnaire (Johnson & Johnson, 1983)	Rattigan, 1997
	Pictorial Scale of Perceived Competence and Social Acceptance (Harter & Pike, 1984)	Robinson et al., 2009
	Partner choice instrumentation (Archer-Kath, Johnson & Johnson., 1994)	Rattigan, 1997
	Adjective Checklist for Describing Classmates (Siperstein, 1980)	Slininger, 1993
	Friendship Activity Scale (Siperstein, 1980).	Slininger, 1993
	Duda and Nicholls' (1992) student satisfaction questionnaire	Viciana et al., 2007
	Several questions researchers used were not from pre-existing questionnaires	Viciana et al., 2007
Psychomotor Domain	Researcher made motor assessment	Alstot, 2011 Slack, 1976
	AAHPERD 1-Mile Run Test (AAHPERD, 1984)	Cramer, 2000
	AAHPERD Basketball Test (AAHPERD, 1984)	Cramer, 2000
	Digi-walker pedometers	Hannon & Ratcliffe, 2004
	Test of Gross Motor Development (TGMD-2; Ulrich 2000)	Martin et al., 2009 Robinson et al., 2011 Valentini, 1991
	Means of scores from Nupponen et al., 1976 assessment	Polvi & Telama, 2000
	On The Ball progress charts, observer evaluations, and set shot, jump stop, and chest pass achievement score (Minneapolis Public Schools, 1988)	Rattigan, 1997
	Basic Motor Ability Test (Arnheim & Sinclair, 1975)	Slack, 1976
	Hamm-Marburg Test (Schilling & Kiphard, 1967).	Slack, 1976
	Neuropsychological test battery used by Dr. R. Trites at the Royal Ottawa Hospital (Not cited)	Slack, 1976
	Polar Heart Rate Monitors	Van Acker et al., 2010
Learning Environment Assessment	TARGET intervention questionnaire (Treasure, 1993)	Cramer, 2000
	Observational Recording Record of Physical Educator's Teaching Behavior (Stewart, 1989)	Hannon & Ratcliffe, 2004
	Teacher-initiated Competitive Orientation scale (TICO; Papaioannou & Kouli, 1999)	Viciana et al., 2007

Article Coding

All studies were coded using a standard coding form (Appendix D). Coding procedures were described in chapter 3. After coding and recoding all articles, two tables were created to provide an overview of the data extracted based on the methodology used, participant information, and study type. Table 4.2 represents all articles used to calculate effect size in the current meta-analysis. This table is divided into three sections which represent methodology, participants, and study information.

Table 4.2: *Table of all studies meeting inclusion criteria*

Study	Methodological												Participant					Study	
	Research Design	Outcome	Design	Time Frame	Manipulation Check	Exercise as Punishment	Establishing the Learning environment	Safety	Diversity	Equity	Cooperation and Competition	Inclusion	School Level	Participant Sex	N	Geographic Location	Type of School	Reporting Method	Publication Type
Alstot, 2011	De	A	Within	0-3 months	No	NR	R	NR	NR	R	NR	NR	E	C	10	USA	NR	Objective	D
Andre et al., 2011	Ex	Mu	Between	0-3 months	Yes	NR	R	NR	R	R	R	NR	M	C	217	France	NR	Combined	J
Balderson, 2006	Ex	A	Between	0-3 months	Yes	NR	R	NR	NR	NR	NR	NR	M	C	6	USA	Public	Combined	D
Barkoukis et al., 2008	De	Mu	Between	>6 months	Yes	R	R	NR	R	R	R	NR	H	C	374	Greece	NR	Self	J
Barkoukis et al., 2010	De	A	Within	>6 months	Yes	NR	R	NR	NR	NR	R	NR	H	C	394	Greece	NR	Self	J
Cramer, 2000	Ex	Mu	Between	4-6 months	Yes	NR	R	NR	NR	NR	R	NR	H	C	65	USA	NR	Combined	D
Escarti et al, 2010A	El	A	Between	>6 months	Yes	NR	R	R	R	R	NR	NR	M	C	42	Spain	Public	Self	J
Goudas & Magotsiou, 2009	Ex	A	Between	0-3 months	Yes	NR	R	NR	R	NR	R	NR	M	C	114	Greece	NR	Self	J
Hannon & Ratcliffe, 2004	De	A	Within	NR	No	NR	NR	NR	R	R	NR	NR	H	C	209	USA	NR	Combined	J
Martin et al., 2009	Ex	P	Between	0-3 months	Yes	NR	R	R	R	NR	R	NR	E	C	64	USA	NR	Objective	J
Polvi & Telma, 2000	Ex	Mu	Between	>6 months	Yes	NR	R	NR	NR	NR	NR	NR	E	F	95	Finland	Comp	Combined	J
Rattigan, 1997	De	A	Between	0-3 months	No	NR	R	R	R	NR	R	NR	E	C	42	USA	Private Head Start Center	Combined	D
Robinson et al., 2009	Ex	A	Between	0-3 months	Yes	NR	R	R	R	NR	NR	NR	E	C	117	USA		Self	J
Robinson et al., 2011	Ex	P	Within	0-3 months	Yes	NR	NR	R	NR	NR	NR	NR	E	C	12	USA	Lab	Objective	J
Slack, 1976	Ex	P	Between	0-3 months	Yes	NR	NR	NR	R	NR	R	NR	E	Ma	19	Canada	NR	Combined	D
Slininger, 1993	Ex	A	Between	>6 months	No	NR	R	NR	R	R	NR	NR	E	C	131	USA	NR	Self	D
Valentini, 1999	Ex	Mu	Between	>6 months	Yes	NR	R	R	R	NR	NR	NR	E	C	67	USA	Early Ed Canter	Combined	D
Van Acker et al., 2010	De	P	Between	0-3 months	Yes	NR	NR	NR	NR	R	NR	NR	M	C	221	Portugal and Belgium	NR	Objective	J
Viciana et al., 2007	Ex	A	Within	0-3 months	Yes	NR	R	NR	NR	NR	R	NR	H	C	95	Spain	NR	Self	J

Note. Ex=Experimental. De=Descriptive. A=Affective. Mu= Multiple. P=Psychomotor. NR=Not reported. R=Reported. E=Elementary school. M=Middle school. H=High school. Comp=Comprehensive School. C=Combination. Ma=Male. F=Female. D=Dissertation. J=Journal article.

The methodology section in table 4.2 included information regarding research design, student outcome (learning domain), design, time frame, manipulation check, and learning environment variables. Eighteen studies used an experimental design, while one used a longitudinal design. Nine studies reported student outcomes in the affective domain, five studies reported student outcomes in multiple domains, five reported outcomes in the psychomotor domain, and zero studies reported learning outcomes in the cognitive domain. Fourteen studies used a between-subjects design, while five used a within-subject design. Eleven studies reported using a 0-3 month time frame, six reported a time frame of greater than 6 months, one reported a time frame of 4-6 months, and one did not report a time frame. Fifteen studies utilized a manipulation check to assess student growth, while four studies did not utilize a manipulation check. Finally, when it came to reporting learning environment variables, 15 studies reported variables relating to establishing the learning environment, 11 studies reported variables relating to diversity, nine studies reported variables relating to cooperation and competition, seven studies reported variables relating to equity, six studies reported variables relating to safety, one study reported variables relating to exercise as punishment, and none of the studies reported variables relating to inclusion.

The second section of table 4.2 included information relating to the study participants. The participant section included data regarding school level, sex, number of participants, geographic location, and school type. Nine studies utilized elementary school students, five studies included high school students, and five studies utilized middle school students. Seventeen studies included a coed population, one study only used female students, and one study only used male students. The number of participants included in each study ranged from 6-394. Ten studies were performed in the United States of America, three in Greece, two in Spain, one in France, one in Canada, one in Portugal and Belgium, and one in Finland. Twelve studies did not report the type of school involved, two reported public schools, one reported a private school, one reported a comprehensive school, one reported a head start center, one reported a lab setting, and one reported an early education center.

Finally, the study section of table 4.2 included information regarding the reporting methods used and the publication type. Seven studies used a combination of objective and self-report methods, seven studies used only self-report measures, and four studies used only objective measures or reporting. Twelve journal articles and seven dissertations were included.

A separate table including all 74 articles included in the final coding process can be found in Appendix E. This additional table also provides a column describing reasons for study exclusion, if any. From the 74 studies that were coded, a sample of 35 studies was re-coded. If there was a discrepancy in the coding, than the study was re-read, and the final interpretation was recorded in both table 4.2, and Appendix E.

Kappa Coefficient

As previously described, a data extraction form was developed by the researcher (Appendix D) to collect data for the meta-analysis. The design for the extraction form was based on (1) methodology used, (2) participant information, (3) study type.

A random sample of 35 articles were re-coded following a 2-month period, which provided the researcher with a kappa coefficient for intra-rater reliability (Table 4.3). Cohen’s kappa (k) is used to measure reliability and adjusts for the chance rate of agreement. Kappa is defined as “the improvement over chance reached by the coders” (Cooper, p.103). Kappa calculations were performed using SPSS software. Based on the results from 595 possible agreements, the kappa coefficient was 0.786 (Table 4.3). According to Viera and Garrett (2005), a kappa coefficient between 0.61-0.80 is considered substantial. Obtaining a 0.786 kappa coefficient established confidence that articles were coded accurately from trial 1 to trial 2.

Table 4.3: *Results of the kappa coefficient calculation using SPSS*

Cohen’s kappa

		Value	Asymp. Std. Error[a]	Approx. T[b]	Approx. Sig.
Measure of Agreement	Kappa	.786	.019	38.861	.000
N of Valid Cases		595			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Effect Size Synthesis

To calculate the effect size synthesis, a summary effect size from each included study was entered into Comprehensive Meta-Analysis (CMA) version-2 software (Borenstein, Hedges, Higgins & Rothstein, 2005). The average treatment effect for all learning environment studies was small (g=0.366; SE=0.111; 95% C.I.=0.163, 0.598; p=0.001) according to Cohen’s (1988) interpretation of effect size as small= 0.2, medium=0.5, and large=0.8. This indicates that the learning environment produced a small effect on student outcomes when compared to no learning environment variables. An effect size of g=0.366 denotes that there is approximately one third of a standard deviation advantage for learning environment treatment groups over control groups.

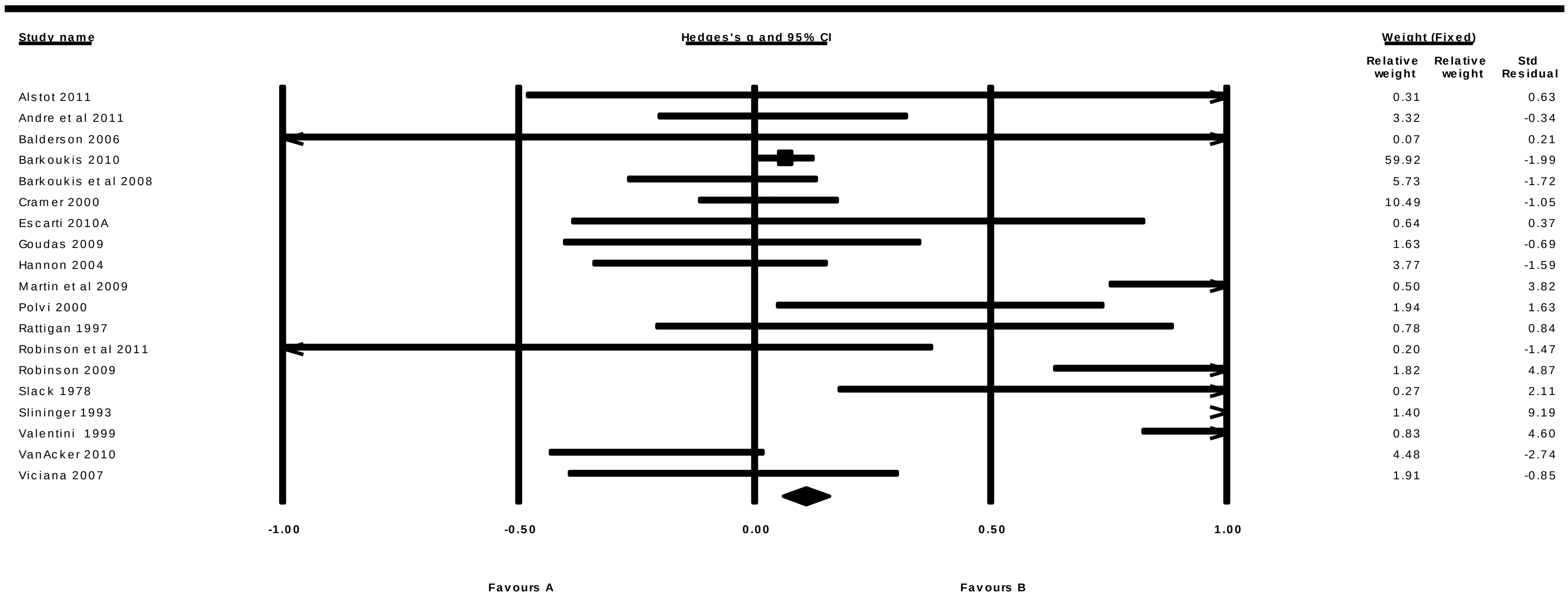
Table 4.4 presents an overview of the relevant statistics used when evaluating the overall effect.

Table 4.4: *Overview of study statistics*

Study name	Statistics for each study						
	Hedges's g	Standard error	Variance	CI Lower limit	CI Upper limit	Z-Value	p-Value
Alstot, 2011	0.385	0.444	0.197	-0.485	1.255	0.867	0.386
Andre et al., 2011	0.060	0.136	0.018	-0.207	0.326	0.439	0.661
Balderson, 2006	0.307	0.963	0.927	-1.580	2.194	0.319	0.750
Barkoukis et al., 2008	0.065	0.032	0.001	0.002	0.128	2.032	0.042
Barkoukis et al., 2010	-0.068	0.103	0.011	-0.271	0.135	-0.656	0.512
Cramer, 2000	0..29	0.076	0.006	-0.120	0.179	0.385	0.700
Escarti et al, 2010A	0.219	0.310	0.096	-0.390	0.828	0.705	0.481
Goudas & Magotsiou, 2009	-0.026	0.194	0.038	-0.406	0.354	-0.136	0.892
Hannon & Ratcliffe, 2004	-0.094	0.128	0.016	-0.344	0.156	-0.737	0.461
Martin et al., 2009	1.433	0.349	0.122	0.749	2.116	4.110	0.000
Polvi & Telma, 2000	0.393	0.178	0.032	0.045	0.742	2.211	0.027
Rattigan, 1997	0.339	0.281	0.079	-0.211	0.889	1.208	0.227
Robinson et al., 2009	-0.703	0.552	0.305	-1.785	0.379	-1.273	0.203
Robinson et al., 2011	0.991	0.184	0.034	0.631	1.351	5.400	0.000
Slack, 1976	1.110	0.477	0.228	0.175	2.045	2.327	0.020
Slininger, 1993	2.017	0.209	0.044	1.606	2.427	9.633	0.000
Valentini, 1999	1.351	0.272	0.074	0.818	1.884	4.970	0.000
Van Acker et al., 2010	-0.208	0.117	0.014	-0.437	0.022	-1.774	0.076
Viciana et al., 2007	-0.045	0.179	0.032	-0.396	0.306	-0.251	0.802
	0.366	**0.111**	**0.012**	**0.163**	**0.598**	**3.428**	**0.001**

Figure 4.2 presents the forest plot representing all included effect sizes. The summary effect size is located at the bottom of the forest plot, and is represented by the large black diamond. An overall effect size to the right of center indicates that there is a positive relationship between learning environment variables and student achievement. The forest plot also highlights that all of the individual effect sizes are reasonably consistent from study to study. Most fall in the range of -0.2 to 0.4, which suggests that it is appropriate to compute a summary effect size.

Figure 4.2: *Forest plot of all studies*



The 95% confidence interval that bounds each effect size is represented by the black horizontal line. If the confidence interval excludes the null (0.0), then the p-value is less than 0.05 and the study is significant (Borenstein et al., 2009). Seven of the studies in the current meta-analysis are statistically significant while 12 are not (Figure 4.2). By scanning the confidence intervals in the forest plot, the statistically significant studies can be identified.

Each individual study's effect size is represented by the black square on the confidence interval line. The square is proportional to the study's sample size in relation to other studies in the meta-analysis (Cooper, 2010). The larger the sample size, the greater the weight of the study in the calculation of overall effect size.

Outlier Analysis and Publication Bias

Following summary effect size calculations, an outlier analyses was necessary due to several outliers found in the final analyses. Since the purpose of a meta-analysis is to provide a reasonable summary of the effect sizes of a body of empirical studies, the presence of outliers may misrepresent the conclusions of a meta-analysis (Viechtbauer & Cheung, 2010). Three outliers were reported (Table 4.5) in the analysis based on their reported z-scores of greater than 1.96. Hannon and Ratcliffe (2004), Slininger (1993), and Viciana et al. (2007) reported z-scores of z=1.99, z=3.66, and z=2.01 respectively. Outliers were controlled by using the "one study removed" function in CMA (Table 4.5). None of the individual studies had a significant impact on the summary effect size if removed.

Table 4.5: *Results of outlier analysis using one study removed function*

	New Summary Hedge's g with Study Removed	Change in Summary Effect size	Lower Confidence Interval	Upper Confidence Interval
Hannon & Ratcliffe (2004)	0.417	-0.036	0.181	0.639
Slininger (1993)	0.249	0.132	0.081	0.418
Viciana et al. (2007)	0.407	-0.026	0.184	0.650
SUMMARY	0.366	-	0.163	0.598

Note. Change in summary effect size was calculated by subtracting the new summary Hedge's g from the original overall summary effect.

After accounting for the presence of outliers, publication bias was addressed. A funnel plot, Trim and Fill procedure, and Fail Safe *N* were analyzed in an effort to address publication bias. Publication bias was deemed moderate as a result of slight asymmetry in the funnel plot (Figure 4.3). However, additional tests established that publication bias was minimal. Zero studies were added during the Trim and Fill procedure. The results of the Trim and Fill (Figure 4.4) procedure indicate that the overall effect size of g=0.366 is unbiased. Results from the Fail Safe N indicate that 212 null studies would be needed for the *p*-value to no longer be significant (Figure 4.5). This large number of studies provides assurance that publication bias is not significant.

Figure 4.3: *Funnel plot of all included studies*

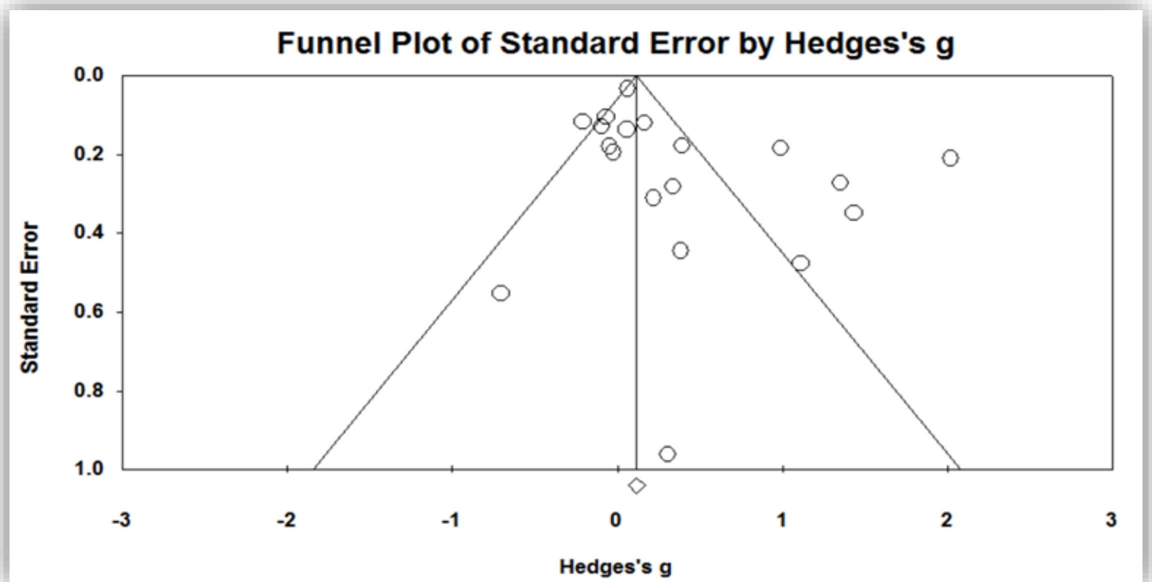


Figure 4.4: *Duval and Tweedie's Trim and Fill*

Duval and Tweedie's trim and fill

		Fixed Effects			Random Effects			Q Value
	Studies Trimmed	Point Estimate	Lower Limit	Upper Limit	Point Estimate	Lower Limit	Upper Limit	
Observed values		0.10533	0.05680	0.15387	0.36590	0.15921	0.57259	168.85981
Adjusted values	0	0.10533	0.05680	0.15387	0.36590	0.15921	0.57259	168.85981

Figure 4.5: *Results from Fail Safe N*

Classic fail-safe N

Z-value for observed studies	6.83186
P-value for observed studies	0.00000
Alpha	0.05000
Tails	2.00000
Z for alpha	1.95996
Number of observed studies	19.00000
Number of missing studies that would bring p-value to > alpha	212.00000

After publication bias was investigated, heterogeneity tests were conducted to determine the variation in the effect sizes (Borenstein et al., 2009). Tests of heterogeneity were calculated using CMA version-2 software (Borenstein, Hedges, Higgins & Rothstein, 2005). Results of the heterogeneity statistics revealed a substantial heterogeneous distribution (Q=167.941, p=0.000, I^2=89.282, T=0.411, T^2=0.169; SE=0.117). The Q statistic was used to determine whether or not there was heterogeneity. A significant Q statistic (Q=167.941, p=0.000) represented heterogeneity in the studies.

After calculating the Q statistic, Tau squared (T^2) and Tau (T) were calculated (T^2=0.169, T=0.411). T^2 represents the variance of the true effect sizes in each study (Borenstein et al., 2009). T^2=0.169 represents a fairly large estimate of the true study variance. T is the estimate of the standard deviation of the true effect size. Based on T=0.411, a relatively large disbursement of effect sizes around the summary effect size.

The final test for heterogeneity was the I^2 statistic (I^2=89.282). The I^2 statistic represents the percentage of variance across studies. Recall that low, medium, and high benchmarks for I^2 are 25%, 50%, and 75% respectively (Higgins, Thompson, Deeks &

Altman, 2003). Since a significant amount of heterogeneity was found, it was necessary to use moderator analyses to explain the variation between the studies.

Moderator Analysis

Moderator analyses were conducted in an attempt to identify features that might explain variability in heterogeneity. Each moderator was considered in a sub analysis with effect size information provided. Table 4.6 presents results from the moderator analyses across all study characteristics.

Methodological characteristics. Significant differences were present in the methodological moderators relating to study design when considering study design (experimental or descriptive) (Q^b=1.04, p<0.05), safety (reported or not reported) (Q^b=5.60, p<0.05), and diversity (reported or not reported) (Q^b=6.47, p<0.05). There were several methodological trends including a.) reporting experimental design (g=0.50, Z=3.59, p<0.05), b.) reporting affective student outcomes (g=0.42, Z=2.36, p<0.05), and c.) reporting a time frame of >6 months (g=0.63, Z=2.70, p<0.05). Studies not employing manipulation checks produced larger effects (g=0.67, Z=2.903, p<0.05) than studies employing checks (g=0.29, Z=2.417, p<0.05). When considering learning environment variables, trends were reported when considering studies that reported establishing the learning environment variables (g=0.46, Z=3.819, p<0.05), diversity variables (g=0.60, Z=4.223, p<0.05), and studies not reporting equity variables (g=0.43, Z=2.851, p<0.05). Overall there was a diverse range of treatment effects, ranging from g=-0.94 to g=0.74.

Table 4.6: *Moderator analyses*

	Effect Size Statistics					Null Test	Heterogeneity Statistics			Publication Bias
	k	*g*	*SE*	s^2	95% C.I.	*Z*	*Q*	τ^2	I^2	Fail Safe N
Random Effects Model [a]	19	0.37	0.105	0.011	(0.159,0.573)	3.470	168.86	0.15	89.34	212
Methodological Characteristics [b]										
Research Design							2.96 [b]			
Descriptive	5	0.05	0.222	0.049	(-0.380, 0.489)	0.24	7.92	0.01	49.49	
Experimental	14	0.50	0.141	0.020	(0.229,0.780)	3.59*	146.78	0.33	91.14	
Outcomes							0.14[b]			
Affective	10	0.42	0.177	0.031	(0.070, 0.763)	2.36*	112.90	0.29	92.03	
Multiple	5	0.31	0.230	0.053	(-0.140,0.760)	1.35	27.43	0.09	85.42	
Psychomotor	4	0.40	0.305	0.093	(-0.202,0.993)	1.30	27.07	0.86	88.92	
Study Design							4.04[b]*			
Between Groups	14	0.52	0.143	0.020	(0.244, 0.805)	3.665*	155.59	0.31	91.65	
Within Groups	5	-0.04	0.242	0.059	(-0.516, 0.434)	-0.170	4.20	0.00	4.77	
Time Frame							2.41[b]			
>6 months	6	0.63	0.235	0.055	(0.174, 1.095	2.698*	111.29	0.34	95.51	
0-3 months	11	0.32	0.191	0.037	(-0.056, 0.695)	0.095	53.53	0.15	81.32	
4-6 months	1	0.30	0.546	0.298	(-1.040, 1.099)	0.957	0	0	0	
NR	1	-0.94	0.555	0.308	(-1.82, 0.994)	0.865	0	0	0	
Manipulation Check							2.22[b]			
No	4	0.67	0.232	0.054	(0.219, 1.130)	2.903*	74.36	1.23	95.97	
Yes	15	0.286	0.118	0.014	(0.054, 0.518)	2.417*	81.14	0.08	82.75	
Exercise as Punishment							1.15[b]			
Not reported	18	.40	0.115	0.013	(0.178, 0.627)	3.512*	165.89	-.017	89.75	
Reported	1	-0.07	0.423	0.179	(-0.896, 0.761)	-0.161	0	0	0.00	
Est. Learning Environment							3.06[b]			
Nor Reported	4	-0.02	0.249	0.062	(-0.511, 0.464)	-0.094	8.35	0.07	64.09	
Reported	15	0.46	0.121	0.015	(0.224, 0.697)	3.819*	151.94	0.17	90.79	
Safety							5.603[b]*			
Not Reported	13	0.22	0.111	0.012	(0.005, 0.439)	2.005*	105.11	0.10	88.58	
Reported	6	0.74	0.187	0.035	(0.370, 1.102)	3.937*	22.07	0.29	77.35	
Diversity							6.469[b]*			
Not Reported	8	0.03	0.172	0.030	(-0.307, 0.368)	0.177	11.65	0.01	39.89	
Reported	11	0.60	0.142	0.020	(0.320, 0.878)	4.223*	137.92	0.45	92.75	
Equity							0.306[b]			
Not Reported	12	0.43	0.149	0.022	(0.133, 0.719)	2.851*	72.43	0.11	84.81	
Reported	7	0.30	0.182	0.033	(-0.060, 0.652)	1.628	96.21	0.36	93.76	

Cooperation & Competition							2.237[b]		
Not Reported	10	0.54	0.155	0.024	(0.231, 0.839)	3.448*	127.85	0.58	92.96
Reported	9	0.21	0.149	0.022	(-0.078, 0.505)	1.436	23.59	0.02	66.09
Inclusion							0[b]		
Not Reported	19	0.37	0.105	0.011	(0.159, 0.573)	3.470*	168.86	0.15	89.34
Participant Characteristics [b]									
School Level							38.24[b]*		
Elementary	9	0.95	0.947	0.126	(0.700, 1.193)	7.527*	55.32	0.43	85.54
High School	5	-0.02	0.177	0.014	(-0.246, 0.213)	-0.141	3.04	0	0
Middle School	5	-0.02	0.148	0.022	(-0.305, 0.275)	-0.103	3.35	0	0
Participant Sex							1.519[b]		
Combined	17	0.34	0.111	0.102	(0,112, 0.558)	3.058*	161.69	0.51	90.10
Female	1	0.39	0.428	0.183	(-0.445, 1.232)	0.919	0	0	0
Male	1	1.11	0.616	0.379	(-0.096, 2.317)	1.804	0	0	0
Geographic Location							7.11[b]		
Canada	1	1.11	0.704	0.495	(-0.269, 2.490)	1.578	0	0	0
Finland	1	0.39	0.547	0.299	(-0.679, 1.465)	0.719	0	0	0
France	1	0.06	0.535	0.286	(-0.989, 1.108)	0.112	0	0	0
Greece	3	-0.01	0.307	0.094	(-0.611, 0.594)	-0.027	1.671	0	0
Portugal and Spain	1	-0.21	0.530	0.281	(-1.247, 0.832)	-0.391	0	0	0
Spain	2	0.07	0.405	0.164	(-0.720, 0.869)	0.183	0.54	0	0
USA	10	0.67	0.193	0.037	(0.293, 1.050)	3.477*	132.02	0.55	93.18
Type of School							10.93[b]		
Comprehensive	1	0.39	0.390	0.152	(-0.371, 1.158)	1.008	0	0	1
Early Ed Center	1	1.35	0.441	0.194	(0.487, 2.215)	3.064*	0	0	1
Head Start	1	0.99	0.393	0.154	(0.222, 1.761)	2.524*	0	0	1
Lab	1	-0.70	0.652	0.425	(-1.981, 0.575)	-1.078	0	0	1
Nor Reported	12	0.28	0.116	0.013	(0.048, 0.501)	2.379*	117.03	0.12	90.601
Private	1	0.34	0.446	0.199	(-0.563, 1.214)	0.759	0	0	0
Public	2	0.23	0.424	0.180	(-0.597, 1.065)	0.552	0.01	0	0
Study Characteristics [b]									
Reporting Method							0.260[b]		
Combined	8	0.38	0.198	0.039	(-0.009, 0.765)	1.913	32.23	0.10	78.28
Objective	4	0.25	0.300	0.090	(-0.336, 0.840)	0.840	22.56	0.67	86.52
Self	7	0.43	0.195	0.038	(0.053, 0.816)	2.232*	112.29	0.28	94.66
Status							7.719[b]*		
Dissertation	7	0.84	0.201	0.041	(0.442, 1.232)	4.158*	97.07	0.91	93.82
Journal	12	0.17	0.133	0.112	(-0.094, 0.427)	1.254	55.68	0.06	80.24

Note. *k* = number of effect sizes. *g* = effect size (Hedges g). *SE* = standard error. S2 = variance. 95% *C. I.* = confidence intervals (lower limit, upper limit). *Z* = test of null hypothesis. τ^2 = between study variance in random effects model. I^2= total variance explained by moderator. * indicates p < .05. a = Total *Q*-value used to determine heterogeneity. b = Between *Q*-value used to determine significance ($\alpha < 0.05$).

Participant characteristics. A significant difference was present in the participant characteristics at the school level category (Q^b=38.24, p=0.05). Students at the elementary school level (g=0.95, Z=7.527, p<0.05) experienced a significant treatment effect, while students in the middle (g=-0.02, Z=-0.103, p>0.05) and high schools (g=-0.02, Z=-0.141, p>0.05) did not. Studies conducted in the USA reported the largest treatment effects of (g=0.67, Z=3.477, p<0.05).

Study features. The moderator analysis found a significant difference between articles published in journals, and unpublished doctoral dissertations (Q^b=7.719, p<0.01). Unpublished doctoral dissertations reported greater treatment effects (g=0.84, Z=4.158, p<0.05) than published journal articles (g=0.17, Z=1.254, p>0.05). Results from reporting methods determined that studies reporting self-assessments had larger treatment effects (g=0.43, Z=2.232, p<0.05) than those reporting objective assessments (g=0.25, Z=0.84, P>0.05) or a combination of assessments (g=0.38, Z=1.913, p>0.05).

Outcome Analysis

Outcome analyses (Table 4.6) generated positive and negative effects ranging from as low as g=-0.87 to g=0.86. No procedures were used to combine similar outcomes, which results in some outcomes including only one effect size. CMA software requires a minimum of three effect sizes to calculate Fail Safe *N*, therefore outcomes that reported less than three effect sizes did not have a corresponding value for N. The largest positive treatment effects for affective outcomes were found for social support (k=2, g=0.75) and perceived competence (k=2, g=0.41). The largest negative treatment effects for affective outcomes were found for negative affect (k=1, g=-0.87) and mastery climate (k=3, g=-0.17). The largest positive treatment effects for psychomotor outcomes were found for locomotor skills (k=4, g=0.86) and skill related fitness (k=3, g=0.75). One negative treatment effect was found for physical activity (k=3, g=-0.14).

Table 4.7: *Outcome analyses*

	Effect Size Statistics					Null Test	Heterogeneity Statistics			Publication Bias
	k	*g*	*SE*	s^2	95% C.I.	*Z*	*Q*	τ^2	I^2	Fail Safe N
Random Effects Model [a]	19	0.37	0.105	0.011	(0.159,0.573)	3.470	168.86	0.15	89.34	212
Affective Outcomes										
Anxiety	1	0.13	0.103	0.011	(-0.070, 0.335)	1.282	0	0	0	-
Attitude	4	0.35	0.539	0.290	(-0.711, 1.401)	0.641	96.77	1.12	96.90	10
Boredom	2	0.21	0.278	0.077	(-0.332, 0.759)	0.767	9.455	0.01	89.42	-
Effort	3	0.14	0.141	0.020	(-0.138, 0.414)	0.980	4.04	0.03	50.63	0
Ego Orientation	3	0.13	0.030	0.001	(0.074, 0.192)	4.424*	1.57	0	0	6
Enjoyment	3	-0.11	0.162	0.026	(-0.422, 0.211)	-0.653	16.94	0.07	88.19	0
Mastery Climate	3	-0.17	0.204	0.041	(-0.571, 0.227)	-0.845	27.05	0.11	92.61	0
Negative Affect	1	-0.87	0.196	0.038	(-1.248, -0.481)	-4.421*	0	0	0	-
Perceived Competence	2	0.41	0.572	0.328	(-0.713, 1.531)	0.714	29.55	0.63	96.62	-
Perception Ability	1	0.22	0.177	0.031	(-0.130, 0.563)	1.225	0	0	0	-
Performance Climate	1	-0.05	0.032	0.001	(-0.11, 0.01)	-1.581	0	0	0	-
Positive Affect	4	0.36	0.203	0.041	(-0.034, 0.760)	1.792	10.72	0.11	72.02	7
Social Skills	5	0.36	0.159	0.025	(0.055, 0.677)	2.308*	9.76	0.07	59.01	12
Social Support	2	0.75	0.181	0.033	(0.396, 1.105)	4.146*	1.30	0.02	22.93	-
Task Orientation	3	-0.05	0.093	0.009	(-0.237, 0.130)	-0.573	5.77	0.02	65.34	0
Psychomotor Outcomes										
Health Related Fitness	2	0.20	0.359	0.129	(-0.509, 0.898)	0.543	14.46	0.24	92.94	-
Locomotor Skills	4	0.86	0.742	0.550	(-.0590, 2.318)	1.165	70.94	2.06	95.77	21
Object Manipulation Skills	6	0.40	0.239	0.057	(-0.073, 0.865)	1.655	25.34	0.25	80.27	19
Physical activity	3	-0.14	0.085	0.007	(-3.06, 0.026)	-1.656	1.28	0	0	0
Skill Related Fitness	3	0.75	0.260	0.067	(0.239, 1.257)	2.879*	4.84	0.12	58.67	15

Note. *k* = number of effect sizes. *g* = effect size (Hedges g). *SE* = standard error. s^2 = variance. 95% *C. I.* = confidence intervals (lower limit, upper limit). *Z* = test of null hypothesis. τ^2 = between study variance in random effects model. I^2 = total variance explained by moderator. * indicates p < .05. a = Total *Q*-value used to determine heterogeneity.

Chapter 5
Discussion

Introduction

This chapter presents a brief summary of the results and study findings based on the research question. Discussions of implications for instructional practice, policy making, and future research are also included.

Summary

The purpose of the current study was to investigate the effect of the physical education learning environment on student outcomes. A total of 19 primary studies were included based on a database search and inclusion criteria. Results of this meta-analysis found that manipulating learning environment variables in physical education had small treatment effects for student outcomes, based on a pooled effect of g=0.366.

Discussion of Findings

The original purpose of the Appropriate Practices documents was to publish a list of appropriate and inappropriate practices for physical educators to consider, based on conventional wisdom and expert opinion. These consensus-based documents have been used throughout the profession to inform many educational decisions. Findings from this meta-analysis indicate that physical education learning environment treatments have a small positive impact on student outcomes. Overall, students involved in learning environment treatments only faired one-third of a standard deviation better than students not involved in treatments. Results from this meta-analysis have significant ramifications for instructional practice, policy, and research.

Instructional practice. The focus of learning environment interventions were based on the seven sub-categories provided in the Appropriate Practices documents.

Results indicate that there is small support to link learning environment variables to student outcomes.

Exercise as punishment. According to the Appropriate Practices documents, exercise as punishment refers to teachers promoting healthy lifestyles and encouraging exercise outside of physical education (NASPE 2009a, 2009b, 2009c; SHAPE, 2009). Teachers do not use exercise as a consequence. In the current meta-analysis, exercise as punishment was only reported in one out of the eighteen studies. Consequently, there is not enough data provided to determine the effect that exercise as punishment has on student outcomes. Subsequent learning environment interventions are needed to better understand this relationship. Current research does not support that using alternative forms of punishment leads to student outcome gains.

Establishing the learning environment. When physical educators establish the learning environment, they actively plan lessons that focus on maximum learning and participation for all children. Children experience a positive learning environment where they feel supported and able to make mistakes (NASPE 2009a, 2009b, 2009c; SHAPE, 2009). Students do not experience harassment from teachers or other students and are provided consistent classroom-management practices where specific behavior expectations are established and consequences are in place. Students develop positive self-concept and are encouraged to focus on intrinsic incentives (NASPE 2009a, 2009b, 2009c; SHAPE, 2009). Variables related to establishing the learning environment were reported in fifteen of the nineteen included studies. The findings support a small positive effect (g=0.46) for treatment groups exposed to variables related to establishing the learning environment.

Safety. According to the AP document, the learning environment is considered safe when teachers provide developmentally appropriate instruction, maintain equipment in excellent condition, and keep facilities free from hazards. Teachers ensure student safety by staying up-to-date on their certifications in first aid, CPR, and AED; by having class sizes consistent with other teachers at their school; by specifically educating students about safety; by practicing emergency action plans; and by monitoring their classes closely (NASPE 2009a, 2009b, 2009c; SHAPE, 2009). A significant difference was found within the safety category. Studies that did and did not report safety interventions resulted in small positive student outcomes. Studies that reported safety variables resulted in higher student outcomes (g=0.74) than studies not reporting safety variables (g=0.22). All studies reporting safety as a variable did so by reporting that the teacher monitored the class. Results indicate that teacher monitoring the class can lead to student outcome gains in the affective and psychomotor domains.

Diversity. The positive learning environment is established when the physical education teacher plans and delivers instruction that includes all children, regardless of race, ethnicity, gender, sexual orientation, physical ability, or religion. The teacher selects activities that represent the interests of the varied cultures of the school community. Differences are acknowledged and appreciated (NASPE 2009a, 2009b, 2009c; SHAPE, 2009). When students were exposed to conditions relating to diversity, they experienced moderate sized gains in outcomes (g=0.60). This finding indicated that controlling for elements of diversity in physical education classes resulted in positive student outcomes.

Equity. Equity in the learning environment refers to the equal inclusion all children regardless of skill level. An equitable learning environment is one where students are encouraged to participate in all activities and are not socialized into activities stereotypically identified as “for boys” or “for girls.” Teachers also promote equity by using gender neutral language (NASPE 2009a, 2009b, 2009c; SHAPE, 2009). Studies not reporting equity variables actually found larger positive effects than studies reporting equity variables. Part of these findings related to the definition of equity in the Appropriate Practices document. No studies reported teachers stereotyping students into activities or using gender neutral language. More studies need to be conducted that control for variables of equity according to the definition in the Appropriate Practices documents if physical educators wish to claim that providing an equitable environment is important for student learning outcomes.

Cooperation and competition. Competition and cooperation refers to the provision of mastery-learning environments where the focus is not always on competition. Students are encouraged to understand different kinds of competition, but also encouraged to set individual goals (NASPE 2009a, 2009b, 2009c; SHAPE, 2009). Similar to variables of equity, it was found that when studies did not report variables of cooperation and competition, students experienced greater learning outcomes. Results indicated again that further studies are needed to support the use of cooperation and competition variables in physical education classes.

Inclusion. In an inclusive learning environment the teacher identifies students with disabilities and follows appropriate accommodation plans for their success. The teacher also accommodates students with temporary medical limitations and differing levels of fitness and skill abilities (NASPE 2009a, 2009b, 2009c; SHAPE, 2009). None of the nineteen studies included learning environment interventions dealing with inclusion. This represents a significant gap in the literature. Although conventional wisdom supports inclusion as an important component of the physical education learning environment, the studies included in the current meta-analysis do not support that an inclusive learning environment has an impact on student outcomes.

Summary. The results for the various treatments differ greatly from each other, implying that some interventions had more positive effects on student outcomes than others. In treatments that examine the effects of safety and diversity interventions with students, the results are significantly more positive than the others. Teachers concerned with using research-based strategies to enhance their teaching should first consider addressing safety and diversity.

Policy-making. These results have important implications for policy makers. NASPE's physical education teacher evaluation tool (1997), for example, provides guidelines for administrators to conduct observations of physical education teachers using similar language from the Appropriate Practices document. In fact, the physical education teacher evaluation tool cites the Appropriate Practice documents as resources for additional information. There is an assumption that providing a positive learning environment leads to student outcomes, and that "quality physical education requires appropriate infrastructure (opportunity to learn)" (NASPE, 2007, p.1). Unfortunately, there is little evidence in the literature to support that learning environment treatments provide enhanced learning opportunities. On the contrary, research does not support that the learning environment has a significant effect on student outcomes. Rather, this meta-analysis indicated only a small positive effect was found on student outcomes. Professionals in the field have made assumptions that cannot be substantiated by the literature, and observation tools were created based on a consensus-based document, rather than a research-based document.

Research. Finally, the results of this meta-analysis have important implications for future research. Several considerations in study methodology should be made when organizing future learning environment research. First, it is imperative that definitions of learning environment are consistent. Language from the Appropriate Practices document should be utilized so that studies can explicitly support the use of the document. It is important that individuals in the physical education field use common language when reporting information. The current phrase "learning environment" had various definitions across the discipline, which made it difficult to find consistency in results.

Another consideration for future research is the use of valid and reliable assessment tools for reporting student data. It is imperative that valid and reliable measures be used when assessing student outcomes to ensure the data gathered is credible and authentic. Assessments lacking validity may not measure what the authors intend. Unreliable assessments may not produce consistent results over time, which makes interpretation of the results extremely difficult. Before research results may be generalized across populations, researchers must be sure that the measures used to assess behaviors are valid and reliable.

There was a clear gap in the literature concerning the exercise as punishment and inclusion learning environment categories. Future studies are especially needed to determine what impact, if any, these learning environment variables have on student outcomes. It is also clear that there is a substantial need for data relating to how the physical education learning environment leads to outcomes in the cognitive domain.

None of the nineteen studies included in the meta-analysis reported outcomes related to the cognitive domain. It is essential that researchers in physical education report student outcomes in all domains since that is what distinguishes physical educators from physical activity leaders. To be an effective physical education teacher, practitioners should know the research-based practices that provide students with the largest student outcome gains.

Attention should also be given to participant characteristics, specifically the school level variable. Results from this meta-analysis showed a significant treatment effect on elementary students ($g=0.95$, $Z=7.527$, $p<0.05$) and negative effects for both middle and high school students ($g=-0.02$). There is a definite need to investigate why learning environment treatments in both middle and high school resulted in negative effects for student outcomes. Common sense suggests that practitioners should utilize the appropriate instructional strategies outlines in these documents, however the results of this meta-analysis indicated negative effects for middle and high school students for establishing the learning environment. Consequently, future research should consider grade level differences for these types of practices.

Limitations of the Study

This meta-analysis faced several limitations that are common in meta-analysis studies. First, one may not conclude that the keyword search of the database represented an exhaustive search of current research articles. The search for articles was limited to five databases, which may not account for all published or unpublished articles on the topic of learning environment and student outcomes in physical education. These databases were selected based on the prominence of physical education research being included.

Additionally, the inclusion criteria in this meta-analysis did not discriminate for the validity and reliability of the assessment tools used to measure student outcomes for each included study. Thirty-two of the thirty-six assessment tools utilized were previously developed, while three were not. Not all studies using previously developed assessment tools reported validity and reliability of the tools. The coding scheme adopted in this meta-analysis allowed for variance in the quality of the assessments in the primary studies.

Further, study delimitations were made in advance. Keywords may have been omitted from search criteria, and consequently articles may have been omitted that would have added to the validity of the findings of this project. Human error may occur when determining what articles should be kept in and which ones should be left out.

Finally, involving only one reviewer to select the final primary studies also created a threat to validity. Although intra-rater reliability was assessed and coding forms were utilized to increase objectivity, some articles may have been coded erroneously or prematurely eliminated due to human error.

Implications for Future Research

It is evident that additional research is needed on this topic. Quantitative research in from the past four decades provided minimal evidence that learning environment treatments lead to student outcomes. It is imperative that the tools being used to educate K-12 students are continuously evaluated for effectiveness, and that policies are in place that use strong, quality data as their foundation. Future research should not be limited to the physical education learning environment, but to all categories included in the Appropriate Practices document.

Appendices

Appendix A

SHAPE America Appropriate Practices Comparison Document

Appropriate Instructional Practice Guidelines, K-12: A Side-by-Side Comparison

SHAPE America – Society of Health and Physical Educators

The following grid includes developmentally appropriate and inappropriate practices in elementary, middle and high school physical education classes. The grid organizes the practices into five separate sections:

1. Learning Environment
2. Instructional Strategies
3. Curriculum
4. Assessment
5. Professionalism

Each section is broken down into subsections that focus on specific areas of concern in physical education.

Appendix B

Database Recording Table

Database & Date Searched	Search Strategy and limits	Results of Search (# of articles)	Total # of Articles for Database	# of Duplicates in Database	# of Articles Excluded and Reasons	# of Articles Included

Appendix C

In/ Out Form

Physical Education Learning Environment In/Out Form

Author and year		Today's date	
Study ID Number		Reviewer	

Question	Yes	Not Clear	No	Further information:
Does the study include learning environment variables for teachers/students?				
Was the study conducted in a Physical Education setting?				
Does the study include one or more of the outcomes measures related to physical education?				
Is the age group studied younger than the **age 18 yrs**?				
Does the article include quantitative descriptive statistics and/or correlations?				
Was the article published from 1970-present?				
English language?				

IF THE ANSWER TO ANY OF THE ABOVE IS SHADED BOX, EXCLUDE THE STUDY (FROM THIS INITIAL SCREENING)

This study is	**Included**	☐	**Excluded**	☐	**Not sure**	☐
	Details:					
Other information						

Appendix D

Data Coding Form

Lead Author(s):______________________	**Study ID Number:**
Methodological-Research Design: 1. Experimental______ 2. Descriptive_______ 3. Longitudinal________ 4. Other:__________________________	**Methodological-Outcome(s)** 1. Psychomotor____ 2. Cognitive______ 3. Affective_______ 4. Multiple Outcomes____________________
Methodological-Design 1. Between Subject______ 2. Within Subject______ 3. Other______________________	**Methodological-Time Frame** 1. 0-3 months _______ 2. 4-6 months ______ 3. Greater than 6 months____________
Methodological-Manipulation Check 1. Yes_____ 2. No______	**Methodological – Exercise as punishment** 1. Children encouraged to participate in PA outside of school ______ 2. Uses exercise as punishment _________ 3. Not reported _____________________
Methodological – Establishing the learning environment 1. Maximum participation__________ 2. Respect and support atmosphere ________ 3. Developing positive self-concept_________ 4. Free of criticism and harassment__________ 5. Emphasis on intrinsic incentives___________ 6. Fair and consistent management __________ 7. Inappropriate behavior consequences ______ 8. Not reported ____________________	**Methodological – Safety** 1. Teacher teaches safety ______________ 2. Practicing emergency action plans_________ 3. Activity for developmental level__________ 4. Teacher CPR, first aid, AED qualifications____ 5. Facilities and equipment maintained and inspected ___________ 6. Class size consistent with other subject areas __ 7. Teacher monitors class _______________ 8. Not reported _______________________
Methodological – Diversity 1. Inclusive and supportive environment ______ 2. Differences acknowledged and appreciated___________ 3. Culturally Diverse Environment ___________ 4. Not reported ___________________	**Methodological – Equity** 1. All children have equal opportunities to participate ___________ 2. Boys and girls are encouraged and supported toward achievement__________________ 3. Teacher uses gender-neutral language _____ 4. Not reported _________________
Methodological – Cooperation and Competition 1. Different kinds of competition are taught ________ 2. Mastery-learning environment ___________ 3. Acceptance of cooperative and competitive student preferences_____________ 4. Not reported ___________________	**Methodological – Inclusion** 1. Teacher uses IEP or school accommodations ____ 2. Lessons adapted for overweight children ______ 3. Adaptations for temporary medical limitations ________ 4. Not reported ______________________
Participant-School Level 1. Elementary School______ 2. Middle School_____ 3. High School_____ Include mean/SD age______/_______ Age range________-________	**Participant-Sex** 1. Male______Number______ 2. Female_____ Number_______ 3. Combined_______
Participant-Geographic Location (country):	Participant – Type of School 1. Public ______ 2. Private ______________ 3. Combination __________________ 4. Other________________
Study-Reporting Method 1. Self-Report_________ 2. Objective Report________ 3. Combined Reporting________	**Study-Publication Type** 1. Journal Article_____________ 2. Dissertation _____________ 3. Other ____________________

Appendix E

Table of All Moderators from Final 74 Articles

Study	Methodological												Participant					Study		
	Research Design	Outcome	Design	Time Frame	Manipulation Check	Exercise as punishment	Establishing the Learning environment	Safety	Diversity	Equity	Cooperation and Competition	Inclusion	School Level	Sex	N	Geographic Location	Type of School	Reporting Method	Publication Type	Reason Excluded
Alarid, 1976	E	Mu	Between	0-3 months	Yes	NR	NR	NR	NR	NR	NR	NR	E	C	82	USA	Public	Combined	D	Not learning environment
Alexander, 1982	D	A	Within	0-3 months	No	NR	NR	NR	NR	NR	NR	NR	H	M	1	USA	NR	Objective	D	Insufficient data
Alstot, 2011	E	A	Within	0-3 months	No	NR	R	NR	NR	R	NR	NR	E	C	10	USA	NR	Objective	D	
Andre et al., 2011	E	Mu	Between	0-3 months	Yes	NR	R	NR	R	R	R	NR	M	C	217	France	NR	Combined	J	
Ashy, 1985	E	P	Between	0-3 months	No	NR	R	R	NR	NR	NR	NR	E	C	80	USA	NR	Objective	D	insufficient data
Balderson, 2006	E	A	Between	0-3 months	Yes	NR	R	NR	NR	NR	NR	NR	M	C	NR	USA	Public	Combined	D	
Bar Eli & Blumenstein, 2004	E	P	Between	4-6 months	Yes	NR	NR	NR	NR	NR	NR	NR	H	C	79	Israel	NR	Combined	J	Not learning environment
Barkoukis et al., 2008	E	Mu	Between	>6 months	Yes	R	R	NR	R	R	R	NR	H	C	374	Greece	NR	Self	J	
Barkoukis et al., 2010	L	A	Within	>6 months	Yes	NR	R	NR	NR	NR	R	NR	H	C	394	Greece	NR	Self	J	
Barkoukis et al., 2012	D	A	Between	0-3 months	No	NR	R	NR	NR	NR	R	NR	H	C	368	Greece	NR	Self	J	Needs more data
Barrett, 2005	E	P	Within	0-3 months	Yes	NR	R	NR	NR	R	R	NR	M	C	23	USA	Public	Objective	J	Time as outcome
Beard, 2011	E	Mu	Between	0-3 months	Yes	NR	NR	NR	NR	NR	R	NR	H	C	91	USA	NR	Combined	D	insufficient data for n
Bekairi et al., 2006	D	A	Within	0-3 months	No	NR	R	NR	R	NR	NR	NR	H	C	265	Greece	NR	Self	J	correlation
Bertelsen, 2002	E	A	Within	0-3 months	No	NR	R	NR	NR	NR	NR	NR	M	C	604	USA	Public	Combined	D	correlation
Blanchfield, 2002	E	A	Within	>6 months	No	NR	NR	NR	NR	NR	NR	NR	M	C	288	USA	NR	Self	D	Not learning environment
Bryant & Curtner-Smith, 2008	E	Mu	Between	0-3 months	No	NR	NR	NR	NR	NR	NR	NR	M	C	113	USA	NR	Combined	J	Not learning environment
Bryant & Curtner-Smith, 2009	E	Mu	Between	0-3 months	No	NR	NR	NR	NR	NR	NR	NR	E	C	201	USA	NR	Combined	J	Not learning environment
Catchings, 2011	D	A	Between	0-3 months	Yes	NR	NR	NR	NR	NR	NR	NR	H	C	204	USA	NR	Objective	D	Not learning environment
Conkell et al, 1999	E	C	Within	0-3 months	No	NR	NR	NR	NR	NR	NR	NR	H	C	543	USA	NR	Combined	J	Not learning environment
Cramer, 2000	E	Mu	Between	4-6 months	Yes	NR	R	NR	NR	NR	R	NR	H	C	65	USA	NR	Combined	D	
Destani, 2011	D	Mu	Within	4-6 months	No	NR	R	NR	R	R	R	NR	H	C	275	USA	NR	Combined	D	correlation
Dyson, 2011	D	NA	Within	4-6 months	Yes	NR	R	NR	NR	R	R	NR	E	C	47	USA	NR	Combined	J	Time as outcome
Elliott et al., 2003	E	P	Between	NR	No	NR	R	NR	R	R	NR	NR	Teachers at elementary	C	20	USA	NR	Combined	D	no pre-post, TEACHERS, NOT STUDENTS
Escarti et al, 2010B	E	A	Between	>6 months	Yes	NR	R	NR	R	R	NR	NR	H	C	30	Spain	Public	Self	J	needs more data
Escarti et al, 2010A	E	A	Between	>6 months	Yes	NR	R	R	R	R	NR	NR	M	C	42	Spain	Public	Self	J	
Farmer, 1985	D	NA	Within	0-3 months	No	NR	R	R	NR	NR	NR	NR	H	C	39	USA	NR	Objective	D	Time as outcome

Study	Methodological												Participant					Study		
	Research Design	Outcome	Design	Time Frame	Manipulation Check	Exercise as punishment	Establishing the Learning environment	Safety	Diversity	Equity	Cooperation and Competition	Inclusion	School Level	Sex	N	Geographic Location	Type of School	Reporting Method	Publication Type	Reason Excluded
Fittipaldi-Wert, 2007	E	NA	Within	>6 months	No	NR	NR	NR	R	NR	NR	R	E	M	4	USA	NR	Objective	D	Time as outcome
Gabbei, 2002	E	P	Between	0-3 months	Yes	NR	NR	NR	R	R	NR	NR	H	C	113	USA	NR	Combined	D	Insufficient data
Gill, 1976	D	A	Within	0-3 months	Yes	NR	NR	NR	NR	NR	R	NR	COLLE	C	92	USA	NR	Objective	D	Not K-12 students
Goudas & Magotsiou, 2009	E	A	Between	0-3 months	Yes	NR	R	NR	R	NR	R	NR	M	C	114	Greece	NR	Self	J	
Halvari et al., 2011	L	A	Within	>6 months	Yes	NR	NR	NR	NR	NR	R	NR	M & H	C	152	Norway	NR	Self	J	correlation
Hannon & Ratcliffe, 2004	E	A	Within	NR	No	NR	NR	NR	R	R	NR	NR	H	C	209	USA	NR	Combined	J	
Harris & Stockton, 1973	E	C	Between	0-3 months	No	NR	NR	NR	R	NR	NR	NR	E	C	120	USA	Public	Objective	J	insufficient data
Hastie & Saunders, 1991	E	NA	Between	0-3 months	No	NR	NR	R	NR	R	NR	NR	E	C	480	Australia	NR	Objective	J	Time as outcome
Hein & Muur, 2004	D	A	Within	0-3 months	No	NR	NR	NR	NR	NR	NR	NR	M & H	C	944	Estonia	NR	Self	J	Not learning environment
Kalaja et al., 2009	D	Mu	Within	0-3 months	No	NR	R	NR	NR	NR	R	NR	M	C	370	Finland	NR	Combined	J	correlation
Karper & Martinek, 1985	E	P	Within	4-6 months	No	NR	NR	NR	R	NR	R	NR	E	C	50	USA	Lab	Objective	J	insufficient data
Kim, 2006	D	NA	Between	0-3 months	No	NR	NR	R	NR	NR	NR	NR	Teachers at elementary	C	4	USA	NR	Combined	D	Teacher behavior outcomes
Langdon, 2010	D	M	Within	0-3 months	No	NR	R	NR	R	NR	R	NR	H	C	157	USA	NR	Combined	D	correlation
MacGillivary, 1980	E	P	Between	0-3 months	No	NR	NR	NR	NR	NR	NR	NR	H	M	90	Canada	NR	Objective	J	Not Physical Education
Martin, 2001	E	Mu	Within	0-3 months	Yes	NR	R	R	R	NR	R	NR	E	C	NR	USA	NR	Combined	D	insufficient n reported
Martin et al., 2009	E	P	Between	0-3 months	Yes	NR	R	R	R	NR	R	NR	E	C	64	USA	NR	Objective	J	
Martinek & Karper, 1984	E	A	Between	4-6 months	No	NR	R	R	NR	NR	R	NR	E	C	126	USA	Lab	Combined	J	needs more data
Melendez, 2004	D	P	Between	0-3 months	No	NR	NR	NR	NR	NR	NR	NR	Teachers at elementary	C	184	USA	Public	Combined	D	Not learning environment
Mitchell, 1992	D	A	Within	0-3 months	No	NR	R	NR	R	NR	R	NR	M	C	622	USA	NR	Self	D	Duplicate
Mitchell, 1996	D	A	Within	0-3 months	No	NR	R	NR	R	NR	R	NR	M	C	622	USA	NR	Self	J	correlation
Moore, 2013	L	A	Within	>6 months	Yes	NR	R	NR	R	NR	R	NR	M	C	216	USA	NR	Self	D	Needs more data
Ommundsen, 2006	D	A	Within	0-3 months	No	NR	R	NR	NR	NR	R	NR	H	C	273	Norway	NR	Self	J	correlation
Ommundsen & Kvalo, 2007	D	A	Within	0-3 months	No	NR	R	NR	R	NR	R	NR	H	C	194	Norway	NR	Self	J	correlation
Polvi & Telma, 2000	E	Mu	Between	>6 months	Yes	NR	R	NR	NR	NR	NR	NR	E	F	95	Finland	Comprehensive school	Combined	J	
Rattigan, 1997	E	A	Between	0-3 months	No	NR	R	R	R	NR	R	NR	E	C	42	USA	Private	Combined	D	data
Robinson et al., 2009	E	A	Between	0-3 months	Yes	NR	R	R	R	NR	NR	NR	E	C	117	USA	Head Start Center	Self	J	
Robinson et al., 2011	E	P	Within	0-3 months	Yes	NR	NR	R	NR	NR	NR	NR	E	C	12	USA	Lab	Objective	J	Needs more data
Rustad, 2007	Not quantitative, suggests solutions but doesn't work with a student population																			Not quantitative

Study	Methodological												Participant					Study		
	Research Design	Outcome	Design	Time Frame	Manipulation Check	Exercise as punishment	Establishing the Learning environment	Safety	Diversity	Equity	Cooperation and Competition	Inclusion	School Level	Sex	N	Geographic Location	Type of School	Reporting Method	Publication Type	Reason Excluded
Gill, 1976	D	A	Within	0-3 months	Yes	NR	NR	NR	NR	NR	R	NR	COLLE	C	92	USA	NR	Objective	D	Not K-12 students
Goudas & Magotsiou, 2009	E	A	Between	0-3 months	Yes	NR	R	NR	R	NR	R	NR	M	C	114	Greece	NR	Self	J	
Halvari et al., 2011	L	A	Within	>6 months	Yes	NR	NR	NR	NR	NR	R	NR	M & H	C	152	Norway	NR	Self	J	correlation
Hannon & Ratcliffe, 2004	E	A	Within	NR	No	NR	NR	NR	R	R	NR	NR	H	C	209	USA	NR	Combined	J	
Harris & Stockton, 1973	E	C	Between	0-3 months	No	NR	NR	NR	R	NR	NR	NR	E	C	120	USA	Public	Objective	J	insufficient data
Hastie & Saunders, 1991	E	NA	Between	0-3 months	No	NR	NR	R	NR	R	NR	NR	E	C	480	Australia	NR	Objective	J	Time as outcome
Hein & Muur, 2004	D	A	Within	0-3 months	No	NR	NR	NR	NR	NR	NR	NR	M & H	C	944	Estonia	NR	Self	J	Not learning environment
Kalaja et al., 2009	D	Mu	Within	0-3 months	No	NR	R	NR	NR	NR	R	NR	M	C	370	Finland	NR	Combined	J	correlation
Karper & Martinek, 1985	E	P	Within	4-6 months	No	NR	NR	NR	R	NR	R	NR	E	C	50	USA	Lab	Objective	J	insufficient data
Kim, 2006	D	NA	Between	0-3 months	No	NR	NR	R	NR	NR	NR	NR	Teachers at elementary	C	4	USA	NR	Combined	D	Teacher behavior outcomes
Langdon, 2010	D	M	Within	0-3 months	No	NR	R	NR	R	NR	R	NR	H	C	157	USA	NR	Combined	D	correlation
MacGillivary, 1980	E	P	Between	0-3 months	No	NR	NR	NR	NR	NR	NR	NR	H	M	90	Canada	NR	Objective	J	Not Physical Education
Martin, 2001	E	Mu	Within	0-3 months	Yes	NR	R	R	R	NR	R	NR	E	C	NR	USA	NR	Combined	D	insufficient n reported
Martin et al., 2009	E	P	Between	0-3 months	Yes	NR	R	R	R	NR	R	NR	E	C	64	USA	NR	Objective	J	
Martinek & Karper, 1984	E	A	Between	4-6 months	No	NR	R	R	NR	NR	R	NR	E	C	126	USA	Lab	Combined	J	needs more data
Melendez, 2004	D	P	Between	0-3 months	No	NR	NR	NR	NR	NR	NR	NR	Teachers at elementary	C	184	USA	Public	Combined	D	Not learning environment
Mitchell, 1992	D	A	Within	0-3 months	No	NR	R	NR	R	NR	R	NR	M	C	622	USA	NR	Self	D	Duplicate
Mitchell, 1996	D	A	Within	0-3 months	No	NR	R	NR	R	NR	R	NR	M	C	622	USA	NR	Self	J	correlation
Moore, 2013	L	A	Within	>6 months	Yes	NR	R	NR	R	NR	R	NR	M	C	216	USA	NR	Self	D	Needs more data
Ommundsen, 2006	D	A	Within	0-3 months	No	NR	R	NR	NR	NR	R	NR	H	C	273	Norway	NR	Self	J	correlation
Ommundsen & Kvalo, 2007	D	A	Within	0-3 months	No	NR	R	NR	R	NR	R	NR	H	C	194	Norway	NR	Self	J	correlation
Polvi & Telma, 2000	E	Mu	Between	>6 months	Yes	NR	R	NR	NR	NR	NR	NR	E	F	95	Finland	Comprehensive school	Combined	J	

Study	Methodological												Participant					Study		
	Research Design	Outcome	Design	Time Frame	Manipulation Check	Exercise as punishment	Establishing the Learning environment	Safety	Diversity	Equity	Cooperation and Competition	Inclusion	School Level	Sex	N	Geographic Location	Type of School	Reporting Method	Publication Type	Reason Excluded
Rattigan, 1997	E	A	Between	0-3 months	No	NR	R	R	R	NR	R	NR	E	C	42	USA	Private	Combined	D	Needs more data
Robinson et al., 2009	E	A	Between	0-3 months	Yes	NR	R	R	R	NR	NR	NR	E	C	117	USA	Head Start Center	Self	J	
Robinson et al., 2011	E	P	Within	0-3 months	Yes	NR	NR	R	NR	NR	NR	NR	E	C	12	USA	Lab	Objective	J	Needs more data
Rustad, 2007	Not quantitative, suggests solutions but doesn't work with a student population																			Not quantitative
Shang, 1998	D	A	Between	0-3 months	No	NR	R	NR	NR	NR	R	NR	M	C	904	China	NR	Combined	D	correlation
Shen, 2004	D	C	Within	0-3 months	Yes	NR	NR	NR	NR	NR	NR	NR	M	C	177	USA	NR	Combined	D	Duplicate
Shen et al., 2009	D	Mu	Within	4-6 months	Yes	NR	R	NR	R	NR	NR	NR	M	C	253	USA	NR	Combined	J	correlation
Shen & Chen, 2006	D	Mu	Within	0-3 months	Yes	NR	NR	NR	NR	NR	NR	NR	M	C	80	USA	NR	Combined	J	Not learning environment
Shen et al., 2007	D	C	Within	0-3 months	Yes	NR	NR	NR	NR	NR	NR	NR	M	C	177	USA	NR	Combined	J	Not learning environment
Shen et al., 2010	D	A	Within	0-3 months	No	NR	R	NR	NR	NR	NR	NR	H	C	566	USA	Public	Self	J	correlation
Sinclair, 2001	D	A	Within	0-3 months	No	NR	R	NR	NR	NR	R	NR	M	C	213	USA	Public	Self	D	correlation
Slack, 1976	E	Mu	Between	0-3 months	Yes	NR	NR	NR	R	NR	R	NR	E	M	19	Canada	NR	Combined	D	
Slininger, 1993	E	A	Between	>6 months	No	NR	R	NR	R	R	NR	NR	E	C	131	USA	NR	Self	D	
Smith & Karp, 1996	E	A	Within	4-6 months	Yes	NR	NR	NR	NR	NR	R	NR	E	C	8	USA	NR	Self	O	insufficient data
Tan, 2000	D	A	Within	0-3 months	No	NR	R	NR	NR	NR	R	NR	M	C	226	USA	Public	Self	D	correlation
Trouilloud et al., 2006	L	A	Within	>6 months	Yes	NR	R	NR	R	NR	NR	NR	H	C	421	France	NR	Self	J	correlation
Valentini, 1999	E	Mu	Between	>6 months	Yes	NR	R	R	R	NR	NR	NR	E	C	67	USA	Early Ed Canter	Combined	D	
Van Acker et al., 2010	E	NA	Between	0-3 months	Yes	NR	NR	NR	NR	R	NR	NR	M	C	221	Portugal and Belgium	NR	Objective	J	
Viciana et al., 2007	E	A	Within	0-3 months	Yes	NR	R	NR	NR	NR	R	NR	H	C	95	Spain	NR	Self	J	
Wilson et al., 2012	D	A	Within	0-3 months	Yes	NR	R	NR	R	NR	NR	NR	E & M	C	533	Canada	NR	Self	J	Needs more data
Wilson, 2012	E	Mu	Between	0-3 months	Yes	NR	NR	NR	NR	NR	NR	NR	M	C	157	USA	NR	Combined	D	Not learning environment
Xiang & Lee, 2002	D	A	Within	0-3 months	No	NR	R	NR	NR	NR	NR	NR	E, M, H	C	308	USA	Private	Self	J	correlation
Zhang et al., 2012	D	A	Within	0-3 months	No	NR	R	NR	NR	NR	NR	NR	M	C	285	USA	Public	Self	J	correlation

Note. Ex=Experimental. L=Longitudinal. D=Descriptive. A=Affective. Mu= Multiple. P=Psychomotor. NR=Not reported. R=Reported. E=Elementary school. M=Middle school. H=High school. C=Combination. Ma=Male. F=Female. D=Dissertation. J=Journal article

References

References marked with an asterisk indicate studies included in the meta-analysis.

* Alarid, J. M. (1976). *An investigation of the relationship between confidence and learning* (Doctoral dissertation). Retrieved from ProQuest. (7625649)

*Alexander, K. R. (1982). *Behavior analysis of tasks and accountability in physical education* (Doctoral dissertation). Retrieved from ProQuest. (8305292)

*Alstot, A. E. (2011). *The effects of a token economy on selected motor skills in physical education* (Doctoral dissertation). Retrieved from ProQuest. (3464535)

American Alliance for Health, Physical Education, Recreation and Dance (AAHPERD). (1984). *AAHPERD skills test manual: Basketball for boys and girls*. Reston, VA: Author.

*Andre, A., Deneuve, P., & Louvet, B. (2011). Cooperative learning in physical education and acceptance of students with learning disabilities. *Journal of Applied Sport Psychology, 23*, 474-485.

Applefield, J. M., Hubler, R., Moallem, M. (2000). Constructivism in theory and practice: Toward a better understanding. *The High School Journal, 84*(2), 35-53.

Archer-Kath, J., Johnson, D. W., & Johnson, R. T. (1994). Individual versus group feedback in cooperative groups. *Journal of Social Psychology, 134*(5), 681-694.

Arnheim, D. D., & Sinclair, W. A. (1975). *The Clumsy Child*. Saint Louis, MO: C.V. Mosby Co.

*Ashy, M. H. (1985). *Student practice and selected teacher behaviors in elementary physical education* (Doctoral dissertation). Retrieved from ProQuest. (8526364)

*Balderson, D.W. (2006). *The effects of a personal responsibility model on individual student and class wide social behaviors* (Doctoral dissertation). Retrieved from ProQuest. (3226608)

Bandura, A. (1990). *Multidimensional scales of perceived academic efficacy*. Stanford, CA: Stanford University.

Bandura, A. (2001). *Guide for constructing self-efficacy scales* (Revised). Retrieved from http://www.des.emory.edu/mfp/self-efficacy.html

*Bar-Eli, M., & Blumenstein, B. (2004). The effect of extra-curricular mental training with biofeedback on short running performance of adolescent physical education pupils. *European Physical Education Review, 10*(2), 123-134.

Barkoukis, V. (2001). *The effect of motivational climate on motivation, anxiety and learning of track and field tasks* (Doctoral dissertation). Aristotle University of Thessaloniki, Greece.

*Barkoukis, V., Koidou, E., Tsorbatzoudis, H., & Grouios, G. (2012). School and classroom goal structures: Effects on affective responses in physical education. *The Physical Educator, 29,* 211-227.

*Barkoukis, V., Ntoumanis, N., & Thogersen-Ntoumani, C. (2010). Developmental changed in achievement motivation and affect in physical education: Growth trajectories and demographic differences. *Psychology of Sport and Exercise, 11*, 83-90.

*Barkoukis, V., Tsorbatzoudis, H., & Grouios, G. (2008). Manipulation of motivational climate in physical education: Effects of a seven-month intervention. *European Physical Education Review, 4*(3), 367-387.

*Barrett, T. (2005). Effects of cooperative learning on the performance of sixth-grade physical education students. *Journal of Teaching In Physical Education, 24*, 88-102.

*Beard, J. (2011). *The effects of traditional and progressive physical education courses on adolescents' body composition and student satisfaction* (Doctoral dissertation). Retrieved from ProQuest. (3489477)

*Bekiari, A., Kokaridas, D., & Sakellariou, K. (2006). Associations of students' self-reports of their teachers' verbal aggression, intrinsic motivation, and perceptions of reasons for discipline in Greek physical education classes. *Psychological Reports, 98*, 451-461.

Berliner, D. C. (1976). A status report on the study of teacher effectiveness. *Journal of Research in Science Teaching, 13*(4), 369-382.

Bernstein, B. (1970). Education cannot compensate for society. *New Society, 15*(387), 344-47.

*Bertelsen, S. L. (2002). *An analysis of the relationships between teacher behaviors and affective outcomes of students in middle school physical education* (Doctoral dissertation). Retrieved from ProQuest. (3059975)

*Blanchfield, L. J. (2002). *The influence of learning environments on intrinsic motivation in urban middle school physical education students* (Doctoral dissertation). Retrieved from ProQuest. (1412788)

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2005). Comprehensive meta-analysis: A computer program for research synthesis (Version 2.0) [computer software]. Englewood, NJ: Biostat.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. United Kingdom: Wiley.

Bowles, S. (1972). 'Schooling and inequality from generation to generation.' *Journal of Political Economy, 80*(3), S219-S251.

Bredekamp, S. (1992). What is "developmentally appropriate" and why is it important? *Journal of Physical Education, Recreation, & Dance, 63*(6), 31-32.

Bronfenbrenner, U. (1978). The social role of the child in ecological perspective. *Zeitschrift für Soziologie, 7,(*1). Retrieved from http://www.jstor.org/stable/23839567

Brookover, W. B., & Lezotte, L. W. (1979). *Changes in school characteristics coincident with changes in student achievement. Occasional paper no. 17.* East Lansing, Michigan: The Institute for Research on Teaching. Retrieved from http://education.msu.edu/irt/PDFs/OccasionalPapers/op017.pdf

Brophy, J., & Good, T. L. (1984). *Teacher behavior and student achievement.* East Lansing, Michigan: The Institute for Research on Teaching.

*Bryant, L. G., & Curtner-Smith, M. D. (2008). Impact of physical education teacher's disability on elementary pupils' perceptions of effectiveness and learning. *Adapted Physical Education Quarterly, 25*, 118-131.

*Bryant, L. G., & Curtner-Smith, M. D. (2009). Influence of a physical education teacher's disability on middle school pupils' learning and perceptions of teacher competence. *European Physical Education Review, 15*(1), 5-19.

Cantwell, R. H., & Andrews, B. (2002). Cognitive and psychological factors underlying secondary school students' feelings towards group work. *Educational Psychology*, *22*, 75–91.

Carlson, T. B. (1995). We hate gym: Student alienation from physical education. *Journal of Teaching in Physical Education, 14,* 467-477.

*Catchings, M. B. (2011). *The effects of an integrated health and physical education program on student achievement* (Doctoral dissertation). Retrieved from ProQuest. (3488040)

Chepko, S., & Arnold, R. (2000). *Guidelines for physical education programs*. Boston: Allyn & Bacon.

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

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., York, R. L. (1966). *Equality of educational opportunity*. Washington, DC: U.S. Government Printing Office.

*Conkell, C. S., Imwold, C., & Ratliffe, T. (1999). The effects of humor on communicating fitness concepts to high school students. *The Physical Educator*, *56*(1), 8-10.

Cooper, H. M. & Hedges, L. V. (Eds.). (1994). The handbook of research synthesis. New York: The Russell Sage Foundation.

Cooper, H. (2010). *Research synthesis and meta-analysis* (4th ed.). Los Angeles, CA: SAGE Publications.

Cooper Institute. (2007). *Fitnessgram/Activitygram test administration manual* (4th ed.). Champaign IL: Human Kinetics.

Copperman, P. (1978). *The literacy hoax: The decline of reading, writing, and learning in the schools and what we can do about it*. New York: William Morrow and Company.

Costelloe, T. (2014). Giambattista Vico. *The Stanford Encyclopedia of Philosophy*. Retrieved from http://plato.stanford.edu/archives/fall2014/entries/vico/

*Cramer, S. J. (2000). *Effects of classroom motivational climate on student goal orientation, attitude towards physical education, cardiovascular fitness, and basketball shooting skill* (Doctoral dissertation). Retrieved from ProQuest. (9965966)

Cronbach, L. J. (1982). *Designing evaluations of educational and social programs.* San Francisco, CA: Jossey-Bass.

Danielson, C. (2007). *Enhancing professional practice: A framework for teaching* (2nd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.

Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Educational Policy Analysis Archives, 8*, (1). Retrieved from http://epaa.asu.edu/ojs/article/view/392/515

Darst, P. W., Pangrazi, R. P., Sariscany, M. J., & Brusseau, T. (2014). *Dynamic physical education for secondary students* (8th ed.). Boston: Benjamin Cummings.

*Destani, F. (2011). *Relationships among the physical education climate and physical activity in junior high school physical education: A mediational analysis* (Doctoral dissertation). Retrieved from ProQuest. (3468222)

Duda, J., Fox, K., Biddle, S., & Armstrong, N. (1992). Children's achievement goals and beliefs about success in sport. *British Journal of Educational Psychology, 62*, 313–323.

Duda, J. L. & Nicholls, J. G. (1992). Dimensions of achievement motivation in schoolwork and sport. *Journal of Educational Psychology, 84*, 290-299.

Duncan M. J., & Biddle B. J. (1974). *The study of teaching*. New York, NY: Holt, Rienhart, & Winston.

*Dyson, B. P., Linehan, N. R., & Hastie, P. A. (2010). The ecology of cooperative learning in elementary physical education classes. *Journal of Teaching in Physical Education*, *29*, 113-130.

*Elliott, S. M. (2003). *The effects of teachers' attitudes toward inclusion on the practice and success levels of children with and without disabilities in physical education* (Doctoral dissertation). Retrieved from ProQuest. (3097258)

*Escarti, A., Gutierrez, M., Pascual, C., & Llopis, R. (2010A). Implementation of the personal and social responsibility model to improve self-efficacy during physical education classes for primary school children. *International Journal of Psychology and Psychological Therapy, 10*(3), 387-402.

*Escarti, A., Gutierrez, M., Pascual, C., & Martin, D. (2010B). Application of Hellison's teaching personal and social responsibility model in physical education to improve self-efficacy for adolescents at risk of dropping-out of school. *The Spanish Journal of Psychology*, *13*(2), 667-676.

Evertson, C. M., & Harris, A. H. (1992). What we know about managing classrooms. *Educational Leadership, 49(*7), 74-78.

*Farmer, J. T. (1985). *Relationship between specified managerial behaviors and student cognitive engagement, motor engagement, and units of academic learning time in a physical education environment* (Doctoral dissertation). Retrieved from ProQuest (8513371).

*Fittipaldi-Wert, J. (2007). *The use of viral supports for students with autism in inclusive physical education* (Doctoral dissertation). Retrieved from ProQuest. (3273355)

Frederickson, N. L., & Furnham, A. F. (1998). Sociometric-status-group classification of mainstreamed children who have moderate learning disabilities: An investigation of personal and environmental factors. *Journal of Educational Psychology*, *90*, 772–783.

*Gabbei, R. C. (2002). *The behaviors and perceptions of teachers and students in middle school physical education as mediated by gender* (Doctoral dissertation). Retrieved from ProQuest. (3059442).

Gage, N. L. (1977). *The scientific basis of the art of teaching.* New York: Teachers College Press.

Gallo, A. M. (2003). Assessing the affective domain. *Journal of Physical Education, Recreation & Dance*, *74*(4), 44-48.

*Gill, D. L. (1976). *The influence of ability composition on group motor performance and intrapersonal variables in competition* (Doctoral dissertation). Retrieved from ProQuest (7708998)

Good, T. L., Biddle, B. J., & Brophy, J. E. (1983). *Teacher effectiveness: Research findings and policy implications.* Washington, D.C.: National Institute of Education.

*Goudas, M., & Magotsiou, E. (2009). The effects of a cooperative physical education program on students' social skills. *Journal of Applied Sport Psychology*, *21*, 356-364.

Goudas, M., Magotsiou, E., & Hatzigeorgiadis, A. (2009). Psychometric properties of the Greek version of the feelings toward group work questionnaire. *European Journal of Psychological Assessment*, *25*, 204–210.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, *5*, 3-8.

Glass, G. V. (2000). Meta-analysis at 25. Retrieved from http://www.gvglass.info/papers/meta25.html

Graham, G., Holt-Hale, S., & Parker, M. (2012). *Children moving* (9th ed). New York: McGraw-Hill.

*Halvari, H., Skesol, K., & Bagoien, T. E. (2011). Motivational climates, achievement goals, and physical education outcomes: A longitudinal test of achievement goal theory. *Scandinavian Journal of Educational Research*, *55*(1), 79-104.

*Hannon, J.C, & Ratcliffe, T. (2004). Physical activity levels in coeducational and single-gender high school physical education settings. *Journal of Teaching in Physical Education, 24*, 149-164.

*Harris, M. B., & Stockton, S. J. (1973). A comparison of bilingual and monolingual physical education instruction with elementary school students. *The Journal of Education Research, 67*(2), 53-54, 56.

*Hastie, P. A. & Saunders, J. E. (1991). Effects of class size and equipment availability on student involvement in physical education. *The Journal of Experimental Education, 59*(3), 212-224.

Harter, S., & Pike, R., (1984). The pictorial scale of perceived competence and social acceptance for young children. *Child Development, 55*, 1969–1982.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New York, NY: Routledge.

*Hein, V., & Muur, M. (2004). The mediating role of cognitive variables between learning oriented climate and physical activity intention. *International Journal of Sport Psychology, 35*(1), 60-76.

Hellison, D. (2003). *Teaching responsibility through physical activity.* Champaign, IL: Human Kinetics.

Higgins, J., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analysis. *British Medical Journal, 327*, 557-560.

Hopple, C. (2005). *Elementary physical education teaching & assessment: A practical guide*. Champaign, IL: Human Kinetics.

Jensen, E. (1998). *Teaching with the brain in mind.* Alexandria, VA: Association for Supervision and Curriculum Development.

Johnson, D. W., & Johnson, R. T. (1983). *Classroom life questionnaire.* Cooperative Learning Center, University of Minnesota, Minneapolis.

Junttila, N., Voeten, M., Kaukiainen, A., & Vauras, M. (2006). Multisource assessment of children‘s social competence. *Educational and Psychological Measurement, 66*, 876–895.

*Kalaja, S., Jaakkola, T., Watt, A., Liukkonen, J., & Ommundsen, Y. (2009). The associations between seventh grade Finnish students' motivational climate, perceived competence, self-determined motivation, and fundamental movement skills. *European Physical education Review, 15*(3), 315-335.

*Karper, W. B., Martinek, T. J., & Wilkerson, J. D. (1985). Effects of competitive/ non-competitive learning on motor performance of children in mainstream physical education. *American Corrective Therapy Journal*, *39,* 10-15.

*Kim, Y. (2006). *Analysis of different class sizes on decision making processes and teaching behaviors of highly experienced teachers (HETs) and less experienced teachers (LETs)* (Doctoral dissertation). Retrieved from ProQuest. (3225368)

*Langdon, J. L. (2010). *Teaching behaviors, student motivation and achievement in the learning domains of physical education* (Doctoral dissertation). Retrieved from ProQuest. (3402793)

Light, R. J., & Pillemer, D. B. (1984). *Summing up: The science of reviewing research.* Cambridge, MA: Harvard University Press.

Lipsey, M. W., & Wilson, D. B. (2000). *Practical meta-analysis* (Vol. 49). London: Sage Publications.

Lund, J. (1992). Assessment and accountability in secondary physical education. *Quest, 44*, 352-360.

Lund, J., & Kirk, M. (2010). *Performance-based assessment for middle and high school physical education* (2nd ed.). Champaign, IL: Human Kinetics.

*MacGillivary, W. W. (1980). Perceptual style, critical viewing time, sand catching skill. *International Journal of Sport Psychology, 11*, 22-33.

Magotsiou, E., Goudas, M., & Hasandra, M. (2006). Validity and reliability of the Greek version of the Multisource Assessment of Social Competence Scale. *Perceptual and Motor Skills*, *103*, 667–675.

*Martin, E. H. (2001). *The effects of a mastery motivational climate motor skill intervention on student achievement and behavior in a naturalistic physical education setting* (Doctoral dissertation). Retrieved from ProQuest. (3028894)

*Martin, E. H., Rudisill, M. E., & Hastie, P. A. (2009). Motivational climate and fundamental motor skill performance in a naturalistic physical education setting. *Physical Education and Sport Pedagogy*, *14*(3). 227-240.

*Martinek, T. J., & Karper, W. B. (1984). The effects of noncompetitive and competitive instructional climates in teacher expectancy effects in elementary physical education classes. *Journal of Sport Psychology*, *6*, 408-421.

Mathis, C. (2011). *What are supports and barriers to using social constructivism and the Reggio Emilia approach in state funded preschool programs in California?* (Doctoral dissertation). Retrieved from ProQuest. (779637406).

McAuley, E., Duncan, T., & Tammen, V. (1989). Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: A confirmatory factor analysis. *Research Quarterly for Exercise and Sport 60*, 48–58.

*Melendez, J. S. (2004). *The relationship between teachers' attitude toward fitness testing and student performance on the Connecticut physical fitness assessment* (Doctoral dissertation). Retrieved from ProQuest. (3117845)

Melograno, V. J. (2007). Grading and report cards for standards based physical education. *Journal of Physical Education, Recreation & Dance, 78*(6), 45-53.

Mercier, K., & Doolittle, S. (2013). Assessing student achievement in physical education for teacher evaluation. *Journal of Physical Education, Recreation & Dance, 84*(3), 38-42.

Miller, A. (2014). Realism. *The Stanford Encyclopedia of Philosophy.* Retrieved from http://plato.stanford.edu/archives/win2014/entries/realism/

Minneapolis Public Schools. (1988). *On the ball* [Motion Picture]. Minneapolis, MN: Minneapolis Public Schools.

Miskel, C., & Ogawa, R. (1988). Work motivation, job satisfaction, and climate. In N. J. Boyan (Ed.) *Handbook of research on educational administration* (pp. 279-304). New York, NY: Longman Publisher.

*Mitchell, S. A. (1992). *Perceptions of learning environment and intrinsic motivation in physical education: Predictive relationship with achievement goals and perceived ability* (Doctoral dissertation). Retrieved from ProQuest. (9308934)

*Mitchell, S. A. (1996). Relationships between perceived learning environment and intrinsic motivation in middle school physical education. *Journal of Teaching in Physical Education, 15,* 369-383.

Mohnsen, B. (2008). *Teaching middle school physical education.* Champaign, IL: Human Kinetics.

*Moore, E. W. G. (2013). *Examining the longitudinal effects of the PE class' climate on students' goal orientations and intrinsic motivation to be physically active* (Doctoral dissertation). Retrieved from ProQuest. (3559164)

*Morgan, K., & Carpenter, P. (2002). Effects of manipulating the motivational climate in physical education lessons. *European Physical Education Review 8*(3), 207-229.

National Association for Sport and Physical Education (NASPE). (2009a). *Appropriate instructional practice guidelines for elementary school* (3rd ed.) [Brochure]. Reston, VA: Author.

National Association for Sport and Physical Education (NASPE). (2009b). *Appropriate instructional practice guidelines for high school* (3rd ed.) [Brochure]. Reston, VA: Author.

National Association for Sport and Physical Education (NASPE). (2009c). *Appropriate instructional practice guidelines for middle school* (3rd ed.) [Brochure]. Reston, VA: Author.

National Association for Sport and Physical Education (NASPE). (1995). *Appropriate practices for middle school physical education* [Brochure]. Reston, VA: Author.

National Association for Sport and Physical Education (NASPE). (1992). *Developmentally appropriate physical education practices for children* [Brochure]. Reston, VA: Author.

National Association for Sport and Physical Education (NASPE). (1998). *Appropriate practices for high school physical education* [Brochure]. Reston, VA: Author.

National Association for Sport and Physical Education (NASPE). (1992). *Developmentally appropriate physical education practices for children. A position statement of the Council on Physical Education for Children.* Reston, VA: Author.

National Association for Sport and Physical Education (NASPE). (1994). *Developmentally appropriate practice in movement programs for young children ages 3-5. A position statement of the National Association for Sport and Physical Education*. Reston, VA: Author.

National Association for Sport and Physical Education (NASPE). (2007). *Physical education teacher evaluation tool.* Reston, VA: Author.

National Association for the Education of Young Children (NAEYC). (2009). *Developmentally appropriate practices in early childhood programs serving children from birth through age 8*. Retrieved from https://www.naeyc.org/files/naeyc/file/positions/position%20statement%20Web.pdf

National Board for Professional Teaching Standards (NBPTS). (2001). *NBPTS physical education standards.* Arlington, VA. Retrieved from http://boardcertifiedteachers.org/sites/default/files/PE_Original%20Standards.pdf

Neff, L. (n.d.). Lev Vygotsky and social learning theories. *Learning Theories.* Retrieved from https://jan.ucc.nau.edu/lsn/educator/edtech/learningtheorieswebsite/vygotsky.htm

Niebuhr, K. (1995). *The effect of motivation on the relationship of school climate, family environment, and student characteristics to academic achievement*. Paper presented at the Annual Meeting of the Mid-South Educational Research Association in Biloxi, MS.

New York State Education Department (NYSED). (2007). *New York State physical education profile.* Retrieved from http://www.p12.nysed.gov/ciai/pe/

Nupponen, H, Telama, R., & Toe Yli, V.- M. (1979). *Koulun kuntotestistoÈ* [Fitness tests for schools].Liikunnan ja kansanterveyden julkaisuja 19. JyvaÈ skylaÈ : JyvaÈ skylaÈ n yliopisto.

*Ommundsen, Y. (2006). Pupils' self-regulation in physical education: The role of motivational climates and differential achievement goals. *European Physical Education Review, 12*(3), 289-315.

*Ommundsen, Y., & Kvalo, S. E. (2007). Autonomy-mastery, supportive or performance focused? Different teacher behaviours and pupils' outcomes in physical education. *Scandinavian Journal of Educational Research, 51*(4), 385-413.

Oxford, R. (1997). Constructivism: Shape-shifting, substance, and teacher education. *Peabody Journal of Education, 72*(1), 35-66.

Papaioannou, A. (1994). Development of a questionnaire to measure achievement orientations in physical education. *Research Quarterly for Exercise and Sport*, 65, 11–20.

*Polvi, S., & Telma, R. (2000). The use of cooperative learning as a social enhancer in physical education. *Scandinavian Journal of Educational Research, 44*(1), 105-115.

*Rattigan, P. J. C. (1997). *A study of the effects of cooperative, competitive, and individualistic goal structures on skill development, affect, and social integration in physical education classes* (Doctoral dissertation). Retrieved from ProQuest. (9734728)

Roberts, G. C., & Balague, G. (1991). *The development and validation of the perception of success questionnaire*. Paper presented in the FEPSAC Congress, Cologne, Germany.

*Robinson, L. E., Rudisill, M. E., & Goodway, J. D. (2009). Instructional climates in preschool children who are at-risk. Part II. *Research Quarterly for Exercise and Sport, 80*(3), 543-551.

*Robinson, L. E., Rudisill, M. E., Weimar, W. H., Breslin, C. M., Shroyer, J. F., & Morera, M. (2011). Footwear and locomotor skill performance in preschoolers. *Perceptual and Motor Skills, 113*(2), 534-538.

Rosenshine, B. (1971). *Teaching behaviours on student achievement.* London, England: National Foundation for Educational Research.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86*, 638-641.

Rosenthal, D. A., Hoyt, W. T., Ferrin, J. M, Miller, S., & Cohen, N. D. (2006). Advance methods in meta-analytic research: Applications and implications for rehabilitation counseling research. *Rehabilitation Counseling Bulletin, 49*, 234-246.

Rubin, D. B., Stuart, E. A., & Zanutto, E. L. (2004). A potential outcomes view of value-added assessment in education. *Journal of Educational and Behavioral Statistics, 29*(1), 103–116.

*Rustad, S. D. (2007). *Coed versus non-coed physical education classes* (Master's thesis). Retrieved from ProQuest. (EP30594)

Rutter, M. (1983). School effects on pupil progress: Research findings and policy implications. *Child Development, 54*, 1-29.

Rutter, M., Maughan, B., Mortimore, P., Ouston, J., & Smith, A. (1979). *Fifteen thousand hours: Secondary schools and their effects on children.* Cambridge, MA: Harvard University Press.

Ryan, R. (1982). Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of Personality and Social Psychology, 43*, 450–61.

Schiemer, S. (2000). *Assessment strategies for elementary physical education.* Champaign, IL: Human Kinetics.

Severly, L. (1975). Individual differences in helping dispositions. *Journal of Personality Assessment, 39*, 282-291.

Schilling, F., & Kiphard, E. J. (1967). *Der Korperkoordinationstest Fur Kinder.* Hunnckens, Kiphard, Kessellman.

*Shang, I. (1998). *An analysis of the relationship between goal perspectives, perceived learning environment, and intrinsic motivation by skill levels and gender in adolescent boys and girls in Taiwan, Republic of China* (Doctoral dissertation). Retrieved from ProQuest. (9925019)

*Shen, B. (2004). *Modeling the interrelations among knowledge, interests, and learning strategies in physical education* (Doctoral dissertation). Retrieved from ProQuest. (3139133)

*Shen, B., & Chen, A. (2006). Examining the interrelations among knowledge, interests, and learning strategies. *Journal of Teaching in Physical Education, 25*, 182-199.

*Shen, B., Chen, A., & Guan, J. (2007). Using achievement goals and interest to predict learning in physical education. *Journal of Experimental Education, 75*(2), 89-108.

*Shen, B., McCaughtry, N., Martin, J., & Fahlman, M. (2009). Effects of teacher autonomy support and students' autonomous motivation on learning in physical education. *Research Quarterly for Exercise and Sport, 80*(1), 44-53.

*Shen, B., Li, W., Sun, H., & Rukavina, P.B. (2010). The influence of inadequate teacher-to-student social support on amotivation of physical education students. *Journal of Teaching in Physical Education, 29,* 417-432.

*Sinclair, C. (2001). *Dispositional goal orientations, perceptions of the motivational climate and intrinsic motivation in secondary physical education: An interactionist approach* (Doctoral dissertation). Retrieved from ProQuest. (3003438)

Siperstein, G. N. (1980). *Instruments for measuring children's attitudes toward the handicapped.* Unpublished manuscript, University of Massachusetts, Boston.

*Slack, A. J. (1976). *The effects of a co-operative perceptual-motor program on the development of motor skills and social co-operation in the learning disabled child* (Doctoral dissertation). Retrieved from ProQuest. (EC55215)

*Slininger, D. G. (1993). *Attitudes toward children with severe mental retardation in integrated physical education* (Doctoral dissertation). Retrieved from ProQuest. (1355032)

*Smith, B. T., & Karp, G.G. (1997). *The effect of a cooperative learning unit on the social skill enhancement of third grade physical education students.* Paper presented at the Annual Meeting of the American Educational Research Association in Chicago, Il.

Society of Health and Physical Education (SHAPE). (2009). *Appropriate instructional practice guidelines, K-12: A side-by-side comparison.* Retrieved from http://www.shapeamerica.org/standards/guidelines/upload/Appropriate-Instructional-Practices-Grid.pdf.

Society of Health and Physical Education (SHAPE). (2013). *Grade-level outcomes for K-12 physical education*. Reston, VA: Author.

Society of Health and Physical Education (SHAPE). (2010). *PE metrics: Assessing national standards 1–6 in elementary school* (2nd ed.). Champaign, IL: Human Kinetics.

Society of Health and Physical Education (SHAPE). (2015). *The essential components of physical education.* Reston, VA: Author.

South Carolina Physical Education Assessment Program (SCPEAP). (2007). *High school notebook.* Retrieved from http://www.scahperd.org/HS_Notebook_1-3-2007.pdf

Stewart, M. (1989). Observational recording record of physical educator's teaching behavior (ORRPETB). In P. Darst, D. Zakrajsek, & V. Mancini (Eds.), *Analyzing physical education and sport instruction* (pp. 249-259). Champaign, IL: Human Kinetics.

Student Achievement. (n.d.). Retrieved September 28, 2015 from http://toolbox.gpee.org/Student-Achievement.242.0.html

*Tan, C. W. K. (2000). *Alienation, goal perspectives, and achievement behaviors in middle school physical education* (Doctoral dissertation). Retrieved from ProQuest. (9996529)

Tomlinson C. A. (1999). *The differentiated classroom: Responding to the needs of all learners.* Alexandria, VA: Association for Supervision and Curriculum Development.

*Trouilloud, D., Sarrazin, P., Bressoux, P., & Bois, J. (2006). Relation between teachers' early expectations and students' later perceived competence in physical education classes: Autonomy-supportive climate as moderator. *Journal of Educational Psychology, 98*(1), 75-86.

Ulrich, D.A. (2000). *Test of gross motor development* (2nd ed.). Austin, TX: Pro-Ed.

USA, U.S. Department of Health and Human Services, Office of the President's Council on Fitness, Sports & Nutrition. (2015). *Active & healthy for LIFE*: *President's challenge physical activity, nutrition & fitness awards program*. Bloomington, IN.

*Valentini, N. C. (1999). *Mastery motivational climate motor skill intervention: Replication and follow-up* (Doctoral dissertation). Retrieved from ProQuest. (9936721)

*Van Acker, R., Carreiro da Costa, F., De Bourdeaudhuij, I., Cardon, G., & Haerens, L. (2010). Sex equity and physical activity levels in coeducational physical education: exploring the potential of modified game forms. *Physical Education and Sport Pedagogy, 15(2),* 159-173.

*Viciana, J., Cervello, E.M., & Ramirez-Lechuga, J.R. (2007). Effect of manipulating positive and negative feedback on goal orientations, perceived motivational climate, satisfaction, task choice, perception of ability, and attitude toward physical education lessons. *Perceptual and Motor Skills, 105,* 67-82.

Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: The kappa statistic. *Family Medicine, 37*(5), 360-363.

von Glaserfeld, E. (1995). *Radical constructivism: A way of knowing and learning*. Bristol, PA: Falmer Press.

Vygotsky, L.S. (1962). *Thought and Language* (2nd ed.). Kozulin, A. (Ed.). Cambridge, MA: MIT Press.

Vygotsky, L.S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press

Warrick, W.R. (n.d.) *Constructivism: Pre-historical to post-modern*. Retrieved from www.unhas.ac.id/hasbi/LKPP/Hasbi-KBK...SCL/.../**constructivism**.pdf

Whitaker, T. (2004). *What great teachers do differently: Fourteen things that matter the most.* Larchmont, NY: Eye on Education.

*Wilson, A. J., Liu, Y., Keith, S. E., Wilson, A. H., Kermer, L. E., Zumbo, B. D., & Beauchamp, M. R. (2012). Transformational teaching and child psychological needs satisfaction, motivation, and engagement in elementary school physical education. *Sport, Exercise, and Performance Psychology, 1*(4), 215-230.

*Wilson, Z. (2012). *The effects of single-gender classes on students' physical fitness test performances and attitudes* (Doctoral dissertation). Retrieved from ProQuest. (3507790)

Wittrock, M. C. (Ed.). (1986). *Handbook of research on teaching* (3rd ed.) New York: Macmillan.

*Xiang, P., & Lee, A. (2002). Achievement goals, perceived motivational climate, and students' self-reported mastery behaviors. *Research Quarterly for Exercise and Sport*, *73*(1), 58-65.

*Zhang, T., Solmon, M.A., Gao, Z., & Kosma, M. (2011). Promoting school students' physical activity: A social ecological perspective. *Journal of Applied Sport Psychology*, *24*(1), 92-105.