DISTRICT DETERMINED MEASURES AS CRITERIA FOR EDUCATOR EVALUATIONS:
PERCEPTIONS OF PHYSICAL EDUCATORS

A dissertation presented
by
Gary M. Nihan

to
The School of Education

In partial fulfillment of the requirements for the degree of
Doctor of Education

In the field of
Education

College of Professional Studies
Northeastern University
Boston, Massachusetts
July 2015
Abstract

This case study has explored the perceptions of physical education teachers around the use of District Determined Measures (DDM’s) as criteria for evaluating teacher effectiveness. Most recently, the Race to the Top initiative has required that teachers be evaluated using multiple sources of data. In Massachusetts, one of the additional sources of data has included the use of District Determined Measures in order to determine teacher effectiveness as it pertains to increased student learning, growth or achievement.

The intent of this study was to determine what effect, if any, the use of DDM’s influences the pedagogical practices differently than the traditional teacher evaluation model that did not include DDM’s. Over the past one hundred years a plethora of research has been conducted around the evaluation of public school teachers. However, there is minimal evidence existing that supports the efficacy around the use of District Determined Measures to improve teacher performance, specifically in content areas such as physical education.

This study, including sixteen physical educators from one school district, provides a small sample of data that may prove useful to school districts that are beginning the process of developing and implementing DDM’s in physical education as well as other content areas that may include the visual arts, performing arts and health education. Two questions have guided this study. They are: 1) How has the new educator evaluation process that now includes District Determined Measures effected change in the physical educators’ instructional practices as perceived by the teacher and by the administrator? 2) Do physical education teachers perceive any change in their professional relationship due to the new evaluation process that now includes District Determined Measures? Key words: District Determined Measures, Value Added Modeling, Student Impact rating, teacher effectiveness.
Acknowledgements

There are many people that have made this project possible and deserve my recognition. Most importantly, the members of my family who are my most significant support system deserve recognition. Without their support and encouragement, I could not have completed this doctoral journey. To my most amazing wife, Marian, who recognized my ambition to undertake this project and as a result provided unconditional support, thank you and I love you! To my son David, whose willingness to take a risk, was influential in undertaking this effort, thank you and I love you! To my daughter Lianne, who recently completed her M.Ed in Special Education and just completed her first year as a special education teacher, your approach to life was also inspirational. Your work ethic is amazing and I love you as well. To my youngest son, Timmy, you were also a key inspiration in this journey. Your ability to bounce back stronger from setbacks and overcome obstacles in life is truly remarkable! Thank you for your part in this process and I love you too! I hope this effort that I have undertaken, at this late stage in my life, can someday be a source of inspiration for each of you.

I would also like to thank my colleague at Salem State University, Dr. Ann Marie Gallo. Your encouragement, ideas, advice and knowledge of APA writing were a very significant part of my success. I would also like to acknowledge Dr. John Passarini, my third reader who is also another colleague at Salem State University. Dr. Pass, as he is best known by his students, and I go back way too long to mention! However, you have been a role model for some time and I appreciate your support as well.

Finally, I would like to thank three professors in the CPS at Northeastern University that have been especially influential. Dr. Chris Unger is an amazing instructor. His ability to encourage me to think outside the box was influential throughout my course work at
Northeastern. Dr. Peg Dougherty, your guidance in the process of determining a thesis topic is especially appreciated. You were able to provide a direction when I was at a crossroads in the process. I am also grateful for the support of my advisor, Dr. Sara Ewell. You were quick to provide feedback and direction. Your mentoring was a extremely significant aspect of this journey. Thank you for your support.
Dedication

I would like to dedicate this doctoral thesis to my father, Richard A. Nihan, Jr. and my mother-in-law, Marianne Warnick. Both individuals provide strong examples of life-long learning to their families. Despite many obstacles in his life, my father attended the Harvard Extension School for six years. In fact, we both received our undergraduate degrees in May of 1974. His commitment to furthering his education was inspirational.

My mother-in-law, Marianne, also furthered her education at the Harvard Extension School later in life, by receiving a Masters Degree of Arts in Foreign Language. Marianne was also extremely instrumental in my doctoral journey by editing all of my numerous papers, including this thesis. Her expertise in grammar, punctuation and sentence structure allowed me to appear a much better writer than I deserve. Without these two individuals in my life, I am not sure I would be earning such a degree.
Table of Contents

Abstract ............................................................................................................................... 2
Acknowledgements .............................................................................................................. 3
Dedication ............................................................................................................................ 5
Table of Contents ............................................................................................................... 6
Chapter I: Introduction ..................................................................................................... 9
   Problem Statement ........................................................................................................ 9
   Significance of Study .................................................................................................... 11
   Positionality Statement ............................................................................................... 13
   Research Questions ..................................................................................................... 16
   Theoretical Framework ............................................................................................... 16
Chapter II: Literature Review .......................................................................................... 23
   The History of Teacher Supervision and Evaluation in America ............................... 24
   The Pitfalls of Teacher Evaluation ............................................................................ 31
   Teacher Effectiveness ............................................................................................... 32
   Ineffective Teacher Evaluation Systems .................................................................... 35
   Effective Teacher Evaluation Systems ...................................................................... 38
   Leadership and Implementation ................................................................................. 41
   Research on Value Added Measures as a Criteria for Teacher Evaluation .......... 43
   Research On The Use Of VAM In Untested Subjects ............................................. 46
   Conclusions ............................................................................................................... 48
Chapter III: Methodology ............................................................................................... 50
Limitations.................................................................................................................107

Conclusion................................................................................................................108

Practical Implications................................................................................................111

Recommendations for Future Research....................................................................112

References..................................................................................................................114
Chapter I: Introduction

Introduction

Race to the Top (RTTT) is a competitive program sponsored by the United States Department of Education (2010), which committed $4.5 billion to this initiative. It is funded through the American Recovery and Reinvestment Act of 2009. The Race To The Top initiative places a strong emphasis on the teacher evaluation process. States that opted to enter the competition and were funded were required to introduce a standards-based approach to teacher evaluation that includes multiple sources of data including Value Added Measures of student achievement as criteria to determine teacher effectiveness.

Problem Statement

The use of Value Added Measures in the educator evaluation process as criteria to evaluate teachers is controversial (Darling-Hammond, 2010). When used correctly, Value Added Measures (VAM) have the potential to quantify teachers contributions in a school year by comparing their student’s state test scores within a given year to the same student’s progress in the previous school year along with the results of other students in that same grade. The use of Value Added Measures is intended to isolate the contribution each teacher makes in that particular school year toward his or her students’ growth. Several noted scholars of education collaboratively authored Problems with the Use of Student Test Scores to Evaluate Teachers (Baker et al, 2010). The summation of their research concluded that student test results should be included in the teacher evaluation process, but should only be one piece of information used to evaluate effectiveness.

Most importantly, the incorporation of Value Added Measures has been recognized as a standard-based teacher evaluation approach and is considered a desirable method of ensuring an
objective process (Harris, 2012). Doulas Harris (2012) stated that Value Added Measures are positively related to almost all commonly accepted measures of teacher performance that include evaluations by principals and classroom observations. Harris (2012), Danielson and McGreal (2000) have stated that a standards-based evaluation system includes defined performance standards, for teachers as well as students. In addition, a rating system for attaining these performance standards should be explicitly stated through the use of a rubric. Danielson and McGreal (2000) also noted that the evaluators must be trained on how to rate teachers, and frequent classroom visits to observe teacher performance must be included. Moreover, multiple criteria should be incorporated to judge student performance and ultimately teacher effectiveness (Danielson & McGreal, 2000). When properly implemented, a teacher evaluation system as described by Danielson and McGreal (2000) has the potential to overcome many of the flaws of traditional teacher evaluation systems, and to stimulate positive teacher reactions to the system.

As a result of the RTTT requirements, the Massachusetts Department of Elementary and Secondary Education collaborated with the Massachusetts teacher unions to develop an educator evaluation program that meets many of the conditions as stated by Danielson and McGreal (2000). The new Massachusetts Educator Evaluation Program included two separate ratings: a Summative Performance Rating and a Student Impact Rating. The Student Impact Rating requires the district to measure the two forms of student growth. Those subjects or grades that are included in the Massachusetts Comprehensive Assessment System must use this test as one measure. The second Student Impact Rating is a district determined measure, or DDM (MA DESE, “District Determined Measures,” 2012). The teachers of subjects that are not included in the Massachusetts Comprehensive Assessment System, also known as untested subjects and grades, are required to develop and employ two District Determined Measures (DDM’s) as a
method to gauge student growth. The grades include Pre-K through grade two and the subject areas are art, music, health education and physical education (“District Determined Measures,” 2012). Accordingly, the use of DDM’s could be considered integral to teachers who instruct in the untested subjects or grades.

In theory, standards that reflect current descriptions of quality instruction provide teachers with targets for performance that are perceived as meaningful and connected to what they are doing (Danielson, 2000). Well-defined standards provide clear definitions for both teachers and evaluators, reducing the subjective nature of the traditional narrative evaluation. Increased frequency of observations and multiple sources of data increase the accuracy of the evaluation and make the results more credible to teachers, thereby building a trust level between teachers and administrators (Colby, 2001). These procedures suggest that evaluation systems as required by RTTT, have the potential to improve schools and, consequently, student learning.

For the purpose of this study the researcher selected physical education as the untested subject.

The ongoing problem of this RTTT initiative, and the research problem of this study, is that physical education teachers are being asked to incorporate a performance based assessment practice when they have had little experience in this approach to assessment.

The researcher anticipated that a change in grading practices might be problematic because many physical educators often use grading criteria derived from a student’s attitude, effort, and proper attire for class. Incorporating an assessment process that is performance-based may become problematic.

**Significance of the Study**

As stated, many physical educators use grading criteria that is compliance based as opposed to performance based measurements. Melograno (2007) contends that students in a
physical education class rarely benefit from compliance-based assessment practices, as they are difficult to quantify and cannot be identified with a standards-based approach. Nonetheless, many physical educators continue to use this approach (Melograno, 2007). Accordingly, there is potential for resistance from current physical education practitioners the measurement of student growth as an indicator of the practitioners effectiveness.

Because physical education teachers, as well as teachers of other untested subjects and grades, will now be held accountable to demonstrate student growth through the use of DDM’s, the results obtained from the study of this problem may prove significant in their respective fields or grades. Moreover, by incorporating the use of District Determined Measures of student growth, physical educators must now place a priority on the process of physical education, skill and knowledge acquisition, rather than the product – physical activity.

Colby (2001) stated that teacher evaluation should be tightly connected to district priorities and school functions such as school improvement, professional development, and student learning. Subsequently, with the expectation that physical educators and/or teachers of other untested subjects and grades will demonstrate student growth, it may become prudent for district administration to provide high quality content specific professional development in order to provide these teachers with the skills and knowledge they need to be proficient in this process.

As a result, individual schools and school districts may find this study significant because it provides information to help administrators understand how teacher evaluation systems can be used as a tool for improving student learning, achievement or growth. The data collected can help school districts identify their own strengths and weaknesses when implementing a growth oriented teacher evaluation system. Because the District Determined Measures are developed locally, with potential for collaboration between teachers and administration, the conclusions that
will emanate from this study can allow districts to improve the process in the development and use of District Determined Measures as viable evaluative criteria. In addition, this study is potentially the first step in documenting a reaction to change that results from teacher involvement in the evaluation process. Finally, the findings from this study will help school districts improve professional development, induction, and hiring programs to strengthen the process of using relevant, multiple criteria in the educator evaluation program.

**Positionality Statement**

The researcher has viewed this problem of practice through the lens of his professional experience, which includes work as a supervisor of in-service as well as pre-service physical educators. The researcher believes that past teacher evaluation practices are inadequate and do not serve the intended purpose. As a result, it appears that opportunities to improve teacher performance and ultimately student performance have been hindered. The researcher believes that the use of District Determined Measures as evaluation criteria for teachers of untested subjects such as physical education has great potential to improve teacher performance, enhance student performance and provide professional growth in the content area of physical education. This belief is founded on the available literature that contends that the DDM’s can be a “valuable tool that can allow educators to see how students are progressing and which teaching methods are having the greatest impact” (MA DESE, 2014).

While the researcher initiated this study through the lens of his professional experience, it must be noted that this experience is limited within the context of two different school districts. In the past, the researcher was familiar with a teacher evaluation format based upon a biannual pre-observation conference, the observation, a post-observation conference and a written summative report. This would occur twice yearly during a teacher’s every other year evaluation
cycle. However, based upon practical experience and available literature it is the researcher’s belief that past teacher evaluation process has shown minimal influence on improving teacher performance or student learning, achievement or growth.

As a requirement of the “Race to the Top Initiative” (2010) some Massachusetts school districts have followed the DESE implementation guidelines and have identified or developed measures for assessing student learning for educators. However, this may or may not have occurred with all subject areas including the untested subjects such as art, music, health education and physical education. The new Massachusetts evaluation system requires that all educators receive a summative performance rating and a student impact rating that is based on the use of District Determined Measures (Mass DOE, 2013). These two ratings will link between professional practice and student achievement (Mass DOE, 2013). Clearly, the DDMs are intended to play a critical role in the decision process of an educator's Student Impact Rating of high, moderate, or low according to the evidence collected regarding student learning, growth and achievement (Mass DOE, 2013). Eventually, every educator will need data from at least two state or district-wide measures in order for trends and patterns to be identified. Since there is no current statewide student assessment process for physical education, the DDMs are currently the only source of data available to be used in the teacher evaluation process (Mass DOE, 2013). This criterion for measuring teacher effectiveness is better recognized in scholarly studies as a Value Added Measure or VAM (Colby, 2001). While it is somewhat controversial, the researcher believes that the incorporation of this criterion has great potential to be utilized as a tool to improve teacher performance, student learning and professional growth. As indicated, many physical educators are unfamiliar with grading criteria that is based upon measurable standards. Because, there is potential for pushback from current physical education practitioners
the researcher is interested in hearing the reaction and perceptions of the physical educators who have used DDMs as teacher evaluation criteria.

The researcher has conducted this study in a larger suburban school district, recognized for a diverse student population that has high expectations for students and staff alike. This district was chosen because there is a large and diverse physical education staff of approximately twenty fully certified teachers who meet the “highly qualified” status as categorized by the No Child Left Behind Act of 2001.

Based upon preliminary conversations with the district wide administration, the researcher was confident that this study would be approved. However, it was necessary to go through a formal request process and follow all IRB expectations. This study was conducted using a qualitative methodology; the participants were interviewed through the use of open-ended research questions that gathered central themes that elicited further questioning (Creswell, 2013). The potential participants were provided a document explaining the study and the researcher was available to respond confidentially to any questions. Because this research topic may be considered controversial, and therefore it may be difficult to elicit honest responses from participants, the researcher investigated and then incorporated the research practices that will best ensure confidentiality and anonymity to participants. Because participation in the study was voluntary, not every school and level in the district was equally represented. Participants were also provided the opportunity to opt out of this study at any time, although no one took advantage of this option. The researcher had minimal familiarity with the staff in this district, a variable that should be avoided to attain accurate data (Butin, 2010).

Because the results of this qualitative study may prove beneficial in the future utilization of DDMs or VAMs as viable criteria for the evaluation of teachers of untested subjects, the
researcher has a strong desire to produce a valid study that avoids any researcher bias. However, the researcher fully recognizes that he has strong sentiments on the topic of this study and therefore researched best practices to avoid this potential problem.

**Research Questions**

The following research questions are proposed for this study and are based upon the perceptions of physical educators.

1. How has the new educator evaluation process that now includes District Determined Measures effected change in the physical educators’ instructional practices as perceived by the physical education teachers?

2. Do physical education teachers perceive any change in their professional relationship with their administrators due to the new evaluation process that now includes District Determined Measures?

These questions were intended to examine the relationship between the evaluative practices of the new educator evaluation process and changes in teacher behavior as perceived by the physical education teacher. Currently, there is limited documented evidence in Massachusetts to support change in teacher performance due to the educator evaluation process utilizing District Determined Measures.

**Theoretical Framework**

The Race to the Top initiative has called for sweeping changes in public education. The implementation of a common core curriculum with associated common assessments as well as a revised approach to teacher evaluation has been the main elements of change (Race to the Top, 2013). The success of these latest reform efforts remains to be seen. Sarason (1991) contends
that chronic failure of school reform has been less because of the ideas, and more so about the implementation process.

Accordingly, the researcher viewed this study through the lens of change theory. There are numerous theories of change that fit a wide range of contexts. The study selected four theories, which are particularly pertinent to education as a lens through which to view this problem of practice. These theories include: Lewin’s Three Step Change Theory, Lippitt’s Phases of Change Theory, Prochaska and DiClemente’s Change Theory and the Social Cognitive Theory. This paper summarizes these relevant theories of change and ultimately details the most pertinent theory or theories selected for the aforementioned problem of practice.

**Lewin’s Three-Step Change Theory**

According to Lewin (1951), the first step in the process of change is to unfreeze the existing equilibrium or status quo. Lewin (1951) states that unfreezing is necessary to overcome resistance by individuals and also to overcome any group conformity that exists. Unfreezing can occur through three methods. First, increase the driving forces that direct behavior away from the existing equilibrium. Second, decrease the restraining forces that can negatively affect the movement to change. Third, find a combination of the first two methods. Lewin (1951) also describes activities that can assist in the unfreezing. They include: a) motivate participants by preparing them for change, b) build trust and recognition of the need for change, and c) actively participate in recognizing problems that may occur and then collaboratively brainstorm solutions (Burnes, 2004).

Lewin’s second step in the process of behavior change is movement. This step in the process of behavior change is necessary to move the target of change to a new level of equilibrium. Lewin provides three actions that can accomplish this task. They are: a) persuading
the targets of change to agree that the status quo is not beneficial to them and encouraging them to view the problem from a fresh perspective, b) working together to seek out relevant information that may assist in moving the change process in the desired direction and c) soliciting and working to incorporating outside input from well recognized leaders in this change process as a resource (Burnes, 2004).

The third step in Lewin’s three-step model of change is refreezing the new level of equilibrium. Obviously, this will occur only after the desired change has been implemented. This step must occur in order to sustain the change. The change will probably be short lived if refreezing does not occur (Lewin, 1951). Institutionalizing the changes through formal and informal mechanisms such as a new policy or procedure can ensure refreezing. Lewin’s model illustrates the effect of potentially divergent forces collaborating to create enduring permanent change.

*Figure 1. Lewin’s Three-Step Change Theory illustration*

**Lippitt’s Phases of Change Theory**

Lippitt, Watson and Westley (1958) extend Lewin’s Three-Step Change Theory by creating a seven-step theory that places a greater emphasis on the role and responsibility of the change agent than the change itself. Lippitt et al (1958) describe the seven steps as follows: 1) Diagnose the problem, 2) Assess the motivation and capacity for change, 3) Assess the resources
and motivation of the change agent to include the change agent’s commitment to change, power and stamina, 4) Choose progressive change objects. In this step, action plans are designed and strategies are developed, 5) The role of the change agents should be selected and clearly understood by all parties so that expectations are clear. Examples of these roles are: cheerleader, facilitator and expert. 6) Maintain the change. Communication, feedback and group coordination are essential elements in this step of the change process and 7) Gradually terminate the helping relationship. The change agent should gradually withdraw from this over time. This will occur when the change becomes part of the organizational culture (p. 58-59).

Lippitt et al (1958) believe that there is one particular condition that will produce an environment where all stakeholders will more readily accept change. This condition occurs when the change phenomenon is widespread. This belief fits well with the change environment anticipated due to the expectations of Race to the Top.

*Figure 2.* Lippitt’s seven stage models of planned change illustration:
Prochaska and DiClemente’s Change Theory

Contrary to Lippett’s Change Theory, Prochaska and DiClemente place an emphasis on the targets of change, rather than the agents of change (Prochaska & DiClemente, 1983). This model is described as cyclical rather than linear. It describes six stages of the change process that exists within the target of change. The six stages include: 1) the pre-contemplation stage which is a phase where the target has no intention of changing yet and is unaware of or denies the need to change, 2) In the contemplation stage the target has recognized that there may be a problem and change is needed but the process has not yet begun, 3) The preparation stage is when the target begins to focus on what can be done and on developing a plan for change, 4) the action stage is when the target publicly states a desire to change, enlists help to make the change and sets realistic goals for the change, 5) the maintenance stage occurs after the change is in place and the target is attempting to prevent any relapse into old habits or behaviors, 6) Termination is when the change is ingrained and ongoing (Prochaska & DiClemente, 1983). This change theory is considered a spiraling model and takes into account behavioral relapses. Yet, the spiral model suggests that the target learns from their relapses instead of circling around the issue.

Figure 3. Prochaska and DiClemente’s change theory illustration:
Social Cognitive Theory

The social cognitive theory also focuses on the target of change and proposes that behavior change is effected by environmental influences, personal factors and attributes of the behavior itself (Fullan, 2006). In this model, the target of change must possess self-efficacy. Essentially this model of change posits that an individual’s acquisition of knowledge or skill relates directly to observation of others within the context of social interactions, experiences and outside media influences. In other words, people will not learn new behaviors strictly by attempting them, whether their behaviors are successful or failing. Rather, the change of behavior is dependent upon the replication of others. Further motivation to change the behavior is the result of observing that the behaviors being potentially replicated are being rewarded due to this change (Kristonis, 2004).

Figure 4. Social cognitive theory illustration:

Framework for the Study

Each of the change theories presented in this paper was worthy of consideration as a lens in which to view this study. However, the framework chosen for this study combined the application of the Lewin’s three-step model of change with Prochaska and DiClemente’s Change Theory. The rationale to support the use of these two theories is logical. Change is unlikely to occur, and be sustained, unless the change agents and change targets are willing and able to
mutually support the change effort. The Lewin model emphasizes the action of the agents of change, the educational leaders. Prochaska and DiClemente’s Change Theory relates to the target of change as an individual and the intrinsic motivation to change.

Finally, the collective writings of Michael Fullan and John Kotter, among the most renowned scholars and authors writing about change, influenced the selection of the two theoretical frameworks used in this study. These two experts in change phenomenon have collectively authored or co-authored more than fifty books and a plethora of papers on the topic. As a result, these seminal authors are cited in thousands of articles and books on the phenomenon of change. While Kotter views the topic of change from the business perspective, Fullan studies change in the genre of education. However, there are similarities in the beliefs of both authors regarding the process of change. In particular, both scholars cite the significance of collaboration.

As an example, in a paper published by the Centre for Strategic Education Seminar Series, Michael Fullan (2006) contends that the most meritorious change theories are those that cite capacity building as their cornerstone. Similarly, in an article written for the Harvard Business Review, Kotter (1995) emphasizes the need for building capacity through a strong coalition of stakeholders who work collaboratively in the process of change. In this particular context of education reform, capacity building occurs when both the educational leaders and the educators are cognizant of the need to change. However, both entities must be willing to work collaboratively and be committed to the effort. The schematic below represents the necessary relationship of three change theories that will ultimately provide the capacity to create and
sustain meaningful change. To be determined is which of the two theories of change will affect the process most readily or, as this researcher hypothesizes, each theory will play a significant role in the change process.

*Figure 5. Combination of Lewin and Prochaska & DiClemente change theories illustration:*

**Chapter II: Literature Review**

Teachers are considered to be the most influential factor on student achievement (Rivkin, Hanushek, & Kain, 2005). Accordingly, it is incumbent upon school systems to evaluate teachers in order to make decisions about their status and improving their performance (Danielson, 2010). In the past, this process typically included a pre-observation conference, the observation, a post-observation conference and a summative report. However, the status quo of the aforementioned teacher evaluation process has been erased in the school districts of those states that have committed to participate in the U.S. Department of Education Race To The Top (RTTT) grant competition.

One significant requirement of all participants in RTTT is an agreement to evaluate teachers using multiple data sources. Having committed to participate in the RTTT competition, the Massachusetts Department of Elementary and Secondary Education has developed an educator evaluation program that requires all school districts to identify measures for assessing
student learning for all educators in all grades and all subject areas (“District Determined Measures,” 2012). Teachers will receive two separate ratings that include a Summative Performance Rating and a Student Impact Rating. The Student Impact Rating requires the district to determine measures, or DDMs, which can rate a teacher’s impact as high, moderate or low based upon student growth (“District Determined Measures,” 2012). In Massachusetts these DDMs, more commonly referred to as Value Added Measures, will play a significant role in assessing a teacher’s performance by tracking the trends and patterns of two different measures of student results over at least two years (“District Determined Measures,” 2012). Where available, the results of the Massachusetts Comprehensive Assessment System must be used for one of the measures. Therefore, additional measures must be developed at the district level for untested subjects such as art, music, physical education, health education and grades pre K through grade two.

While acknowledging problems with the previous systems of teacher evaluation, the use of Value Added Measures (VAMs) to assess teacher effectiveness has been considered problematic by many educational scholars (Baker et al, 2010). Prior to reviewing the available literature on the use of DDMs or VAMs to evaluate teacher effectiveness, it will be relevant to examine the plethora of literature on teacher evaluation to include: the evolution of teacher evaluation, past and current perspectives on the purpose of teacher evaluation, past and current perceived qualities of effective teaching, views on ineffective and effective teacher evaluation systems, and research on standards based teacher evaluation programs.

The History of Teacher Supervision and Evaluation in America

In the 1700’s, educators did not require any specific professional preparation. The hiring, supervision and evaluation of teachers was typically left in the hands of local government and
local clergy within rural settings (Marzano, Frontier & Livingston, 2011). The clergy were considered a logical choice for this supervisory role due to their level of education and the influence that religion had upon the curriculum at that time (Tracy, 1995, p. 320). During this time, teachers were viewed as servants of the community and as a result had very little input on curriculum and instructional methodology. Supervisor’s had unlimited authority for hiring and dismissal teachers as well as for determining the criteria for effective instruction (Burke & Krey, 2005). With no established criteria for what constituted effective instructional practices or sound pedagogical expertise, the quality and type of feedback was extremely varied during this era (Marzano et al, 2011).

The common school movement and the rise of an industrial based economy became prevalent throughout the 1800’s. This phenomenon in large urban settings generated a need for more complex school systems and a demand for teachers with expertise in specific disciplines as well as administrators who were capable of assuming complex responsibilities. Typically, one teacher within a school was selected to assume these administrative responsibilities. This principal teacher eventually evolved into the building principal. This prevailing trend expanded into rural school districts and became the standard supervisory model (Marzano et al 2011). At this time, it was also recognized that clergy did not possess the expertise to make informed decisions about teacher effectiveness (Tracy, 1995). Tracy explains “Rather than simply understanding the mores of the community, the supervisor now needed to have subject area knowledge and teaching skills” (p. 323). By the mid -1800s, teaching was perceived as a multifaceted profession that required sophisticated feedback in order for instruction to improve. Accordingly, supervision began to focus on improving instruction (Blumberg, 1985). From the onset of formal education in the United States up to the mid - 1800’s there was a realization that
pedagogical skills were a very significant aspect of effective teaching. Although there was little or no formal discussion about the specifics of these skills, the acknowledgment of their importance might be considered the first step in the journey to a comprehensive approach to developing teacher expertise (Marzano et al, 2011).

During the latter half of the nineteenth century and for part of the twentieth century there were two competing views of education. In the early twentieth century, John Dewey promoted a progressive view of education that emphasized democratic ideals, citizenship and progressive ideas such as student centered education that connected the real world with the classroom and differentiated instruction that met individual student needs and interdisciplinary instruction (Dewey, 1938, 1981).

In contrast to John Dewey there was Frederick Taylor. Taylor was a nineteenth century mechanical engineer that espoused a scientific view of industry management and believed that the measurements of specific behaviors of factory workers would likely be the most powerful method to increase worker productivity (Marzano et al 2011). Taylor also believed that most efficient method could be determined by studying the multitude of ways a task could be performed. Furthermore, Taylor believed that this method could be applied to specific tasks such as shoveling coal and more universal tasks such as the selection of workers and the development of training programs. Taylor’s ideas resonated with business owners, engineers and colleges of business and engineering. Eventually, these ideas were accepted and entrenched into K-12 education (Marzano et al, 2011). Eventually, Taylor’s philosophy of scientific measurement became the accepted tool for instilling a more scientific approach to education. Edward Cubberly’s book *Public School Administration* (1929,) presented Taylor’s principles and detailed how they could be used to manage schools the way factories are managed. Cubberly’s book laid
out a set of principles for administrators that emphasized the use of data to inform teachers of their effectiveness based upon a formal observation. By applying a letter grade scale of A through F, Cubberly (1929) was able to rate a teacher’s performance during an observation.

Building upon some of Cubberly’s scientific methods, William Wetzel (1929, as cited in Marzano et al, 2011) was the first recognized advocate for including multiple sources of data to determine teacher effectiveness. These sources included the use of student aptitude tests to determine the ability level of each child; the establishment of clear, measurable objectives for each course, and the teacher’s effective use of specific instructional practices. This era is sometimes perceived as the precursor for the current approach to education.

Throughout the 1930s, there was continued tension between the scientific approach to schooling, which included an emphasis on standardized testing, and the educational approach that emphasized social development and democratic values. However, although Wetzel and Cubberly’s data driven model of schooling conflicted with Dewey’s model, which focused on the ultimate goal of education, there was room for collaboration. As a result a broader perspective of teacher supervision promoting the use of data and valuing the ultimate goal of education evolved during this time.

The post World War II era of teacher evaluation and supervision shifted away from the scientific approach and valued the teacher in a humanistic manner. The teacher was recognized as a person undertaking a challenging and most important role in society. Lewis and Leps (1946) developed guidelines for a supervisory model including: (1) democratic ideals, (2) opportunities for initiative, (3) understanding human limitations, (4) shared decision making, and (5) delegation of responsibility (p. 163). Describing this new approach to teacher supervision, Lewis and Leps (1946) stated, "The school administrator, with the acceptance of the community,
is gaining the courage to utilize the creative force to be gained in freeing the human beings who
comprise the school situation to participate in the making of policies and plans for their
execution; and, hence, to utilize the force and creativity inherent in the democratic process" (p. 161).

While the teacher was treated with a newly found level of respect, the supervisors role
during this era was greatly increased to include the following areas: the curriculum, teaching
personnel, the teaching/learning situation, the emotional quality of the classroom, resources and
materials of instruction, auxiliary functions including working with the school lunch service,
attendance, distribution of textbooks, public relations, and working with cooperative groups and
agencies (Swearingen, 1946). William Melchoir (1950) included additional supervisory
responsibilities in his text *Instructional Supervision: A Guide to Modern Practice*. These
responsibilities included meetings with teachers, faculty meetings, business meetings, social
meetings, workshops and other committee meetings in addition to "classroom visitation for
observation and study" (p. 51). While classroom visitations were specifically included in the
responsibilities, its significance was undervalued when compared with the management of the
building facilities. This is apparent when noting that twenty-three pages were devoted to
“beautifying the grounds” in comparison with only sixteen pages dedicated to classroom visits.
Ethel Thompson (1952) noted that a growing number of administrative responsibilities were
added to supervisors during this era. Thompson (1952) describes these additional administrative
duties as follows; attending student placement conferences; observing in a classroom; working
with parents and other principals; completing paperwork; meeting with various school
committees; attending student conferences; recruiting new teachers; meeting with various
professional organizations; doing demonstration lessons, and acting as a resource to others in the organization.

Even though the explosion of additional supervisory responsibilities was ineffective at best, one positive outcome from this era was the consensus among teachers and administrators that classroom observations were an important and valuable aspect of teacher development. In an article “Teachers Look at Supervision,” Matthew Whitehead (1952) explained six broad areas of supervision and surveyed teachers to determine their perceptions as to the importance of each area. The results from the survey indicated that advances in the process of teacher observation were needed. In particular, Whitehead (1952) noted that for an efficient and worthwhile observation, principals needed to be present for an entire class period and conduct a meaningful post-observation conference. Whitehead (1952) summarized his position by stating that administrators must pay more attention to the primary aim of education: effective teaching.

In the late 1950s, Morris Cogan, a professor in the Master of Arts in Teaching at Harvard University, developed a clinical model of teacher supervision (Marzano et al, 2011). His model was a systematic approach designed for providing purposeful feedback to student teachers with an approach similar to the supervisory practices used in teaching hospitals. Robert Goldhammer, a student of Cogan at Harvard University, further developed his clinical approach which involved “a purposeful, symbiotic relationship between the practitioner and resident, where observation and discussion drove both parties to higher levels of growth and effectiveness” (Goldhammer, 1969, p.54). This revised model of clinical teacher observation incorporated five phases. These included: a) pre-observation conference which was intended to provide a framework for the lesson, b) classroom observation where the supervisor observed the teacher utilizing the framework and collected data, c) analysis of the data collected by the supervisor, d)
post conference with a dialogue between the supervisor and teacher concerning the data collected, e) examination of the analysis where the supervisor’s practice was examined with all the rigor and for the same purposes that the teacher’s professional practice was analyzed.

In 1973, Cogan wrote the book *Clinical supervision*. It was in this text that Cogan illuminated several points regarding teacher supervision that resonate with the today’s RTTT educator evaluation process. Specifically, Cogan (1973) contended that the cornerstone of a supervisor’s work with the teacher is the understanding that the clinical teacher evaluation is a continuation of a teacher’s professional development. Most importantly, Cogan (1973) believed that the teacher must feel that he/she is treated with respect and states, “He must not be treated as a person being rescued from ineptitude, saved from incompetence, or supported in his stumblings. He must perceive himself to be engaged in the supervisory processes as a professional who continues his education and enlarges his competences” (p. 21).

In 1983, Thomas McGreal outlined a range of supervisory practices that essentially provided differentiated supervisory models based upon the teacher’s experience. McGreal (1983) stated that non-tenured teachers and those teachers who have shown significant instructional deficiencies would receive an intensive developmental supervision model. Accordingly, an experienced teacher who has demonstrated solid instructional practices over time would participate in a self-directed professional development plan. This model aligns closely with the practices that are promoted in new RTTT educator evaluation program.

In the 21st century, the Danielson “framework for teaching” is the most widely used model of teacher supervision (Danielson, 1996). This framework for teaching contains four domains of instruction: a) planning and preparation, b) instruction, c) classroom environment and d) professional responsibilities. Accompanying this model is a four level rubric that is intended
to measure instructional proficiency. The levels range from unsatisfactory to distinguished teaching practice. Most importantly, this system is generic and is claimed to apply to all grade levels and content areas (Kimball & Milanowski, 2009). Danielson’s system has been validated and reported to have shown improved student learning. The Massachusetts new educator evaluation system has used the Danielson model as a guide in the design (personal conversation, August 8, 2014, Massachusetts Department of Elementary and Secondary Education).

**The Pitfalls of Teacher Evaluation**

With so much emphasis on improving standardized testing scores and year-to-year student growth, teachers are defined and held accountable for these results (Hinchey, 2010). Using a single measurement in making high stakes decisions such as termination, pay, and tenure, has become a major issue within our public school system. There is evidence that value added measures in teacher evaluation systems may be invalid and unreliable sources in determining teacher quality and student achievement (Amrein-Beardsley & Collins, 2012; Hill & Herlihy, 2011). Furthermore, tying sanctions to test scores and value added measures can potentially yield consequences such as discouragement of teachers wanting to work in underachieving schools, along with demoralization of teaching in general (Rothstein et al., 2010). Marion and Buckley (2011) pointed out that many Race to the Top applications included promises that states would use other forms of data to demonstrate student achievement in non-tested subjects. This poses a potential problem for physical educators based upon lack of national standardized assessment measures to demonstrate student learning, along with the unfair assumption that physical education teachers are sharing responsibility of student achievement on a school-wide level.
Teacher Effectiveness

With the release of the 1983 study called *A Nation at Risk: The Imperative of Education Reform* (National Commission on Excellence in Education, 1983), the U.S. Government became serious about student learning. In particular, teacher accountability was aligned with student learning more so than ever in the past. Mehta (2013) argues that the Nation at Risk report created a domineering example that has shaped school policy since it was released in 1983. As a result educators in the 1980’s and 1990’s were under increasing scrutiny to prepare students for a changing job market. Students were now required to develop a deeper understanding by learning more complex skills such as critical thinking, problem solving, and collaborative learning. Moreover, there was an increased understanding of how children learn. Research on teaching methodology and content developed new insights into how content, teacher, learner, and context must be interconnected for effective teaching and learning to occur (Danielson & McGreal, 2000).

In 1996, the National Commission on Teaching and America’s Future published a report, *What Matters Most: Teaching for America’s Future* (Darling-Hammonds, 1997) which developed a series of recommendations that covered the entire continuum of teacher development, including teacher evaluation standards. The first recommendation linked teacher standards to student standards. It made the connection between what effective teachers should know and be able to do in order to help students succeed at meeting the new standards that will prepare them for a twenty-first century workforce (Darling-Hammond, 1997). These standards were developed collaboratively by three groups: the National Council for the Accreditation of Teacher Education (NCATE), which sets standards for schools of education; the Interstate New Teacher Assessment and Support Consortium (), which develops standards for the licensing of
beginning teachers; and the National Board for Professional Teaching Standards (NBPTS), which sets standards for accomplished practice (Darling-Hammond, 1997). These three different sets of standards are aligned with one another and with teacher assessments. These assessments measure how teachers meet these standards by examining with teachers the evidence of their teaching performance using such artifacts as videotapes of teaching, lesson plans, student work, and analyses of curriculum. The rigorous standards set by these three organizations truly provided an opportunity to use formative evaluation as a means to foster teacher growth and improve student learning.

The “Widget Effect” is a widely read report developed by the New Teacher Project. This 2009 report examines the failure of our nation’s education system to recognize and respond to variations in the effectiveness of our teachers. The report contends that theoretically our teacher evaluation programs should “serve as the primary mechanism for assessing such variations, but in practice tell us little about how one teacher differs from any other, except teachers whose performance is so egregiously poor as to warrant dismissal” (Weisberg et al, 2009). This report begins with a conjured, but potentially realistic scenario that asks: if you were a parent entering a new school district, how would you determine which teacher(s) are recognized as the best other than from parental word of mouth? The question is intriguing since this report cites numerous literature sources that contend that as many as 98% of teachers today are recognized as satisfactory based upon formal observations. This report further argues that it is highly unlikely you would be able to obtain information relative to which teachers are the most effective, least effective or even who falls in the middle of this paradigm. Simply put, in the United States there is currently an inability to differentiate excellent teaching such as excellent from good, good from fair and fair from poor. The report states that school districts tend to assume that teacher
effectiveness varies little from teacher to teacher and as such, teachers are interchangeable parts and not individual professionals with strengths and weaknesses. The report further argues that these beliefs are disrespectful to teachers and ultimately have a negative effect on student achievement.

The research in the “Widget Effect” report was derived from survey responses of twelve demographically diverse school districts in 4 states that included Ohio, Illinois, Arkansas and Colorado. Over fifteen thousand teachers, thirteen hundred administrators and eighty local and state education officials contributed to the data collected. The results of this report are intriguing yet not surprising. Regardless of the range of rating options used to determine teacher effectiveness, only one percent of teachers are rated as unsatisfactory. This means that excellence in teaching performance goes unrecognized and unrewarded. Consequently, when variations in teacher effectiveness go unrecognized there is a strong likelihood that the specific needs of professional development go undetermined. Additionally, this report cited that 66 percent of novice teachers receive ratings of satisfactory or above. This also influences teacher retention at the conclusion of the novice teacher probationary period. In fact, forty-one percent of the surveyed administrators have never “non-renewed” teachers in their final probationary year. Furthermore, eighty-one percent of administrators and fifty-seven percent of teachers cite tenured teachers in their building who are performing poorly. Moreover, forty-three percent of teachers surveyed stated that there is a tenured teacher in their building who should be dismissed. To compound these results, these percentages are even higher in high-poverty districts.

The results of the “Widget Effect” demonstrate a serious concern about the current methods of teacher evaluation. It suggests a need to develop a teacher evaluation system that can provide data that is measurable and can demonstrate tangible evidence of teacher effectiveness.
In response to this concern many policy makers believe that tangible evidence of teacher effectiveness can be determined by calculating the improvement in student tests scores (Danielson, 2011).

Based upon research, Linda Darling-Hammond (2010) has found that more effective teachers possess the following qualities:

1. Strong general intelligence and verbal ability that help them organize and explain ideas, as well as observe and think diagnostically
2. Strong content knowledge in the areas they teach
3. Knowledge of how to teach others in that area (content pedagogy), in particular how to develop higher-order thinking skills
4. An understanding of learners and their learning and development - including how to assess and scaffold learning, how to support students who have learning differences or difficulties and content for those who are not already proficient in the language of instruction.
5. Adaptive expertise that allows them to make judgments about what is likely to work in a given context in response to students’ needs.

Darling – Hammond (2010) summarizes this list as a willingness to support learning for all students.

**Ineffective Teacher Evaluation Systems**

The majority of teacher evaluation systems tend to focus on summative methods, rather than formative methods and the limited criteria used to evaluate them. The systems in practice prior to RTTT, were likely developed 30 years ago and have not been updated to reflect current research on effective teaching practices. These systems rely on the documentation of a small
number of observable behaviors that focus on isolated qualities of classroom management, effective routines and procedures, and teacher personality (Danielson & McGreal, 2000). Annual goal setting and review are often part of this process and some districts even add a form of clinical supervision (Peterson, 2004). While these are useful components of teacher evaluation, they do not adequately foster teacher growth and improvement.

There are other problems associated with past teacher evaluation systems. First, there was a lack of precision in evaluating performances. Most of these systems were not standards-based and relied on a scale that had not been calibrated to identify the different levels of a standard. In addition, the culture in many school environments would not support teacher growth, thereby creating the expectation that all veteran teachers are on the higher end of the rating scale. Shinkfield and Stufflebeam (1995) describe a case study of an urban school district where 95 percent of the teacher evaluations indicated that the teachers were superior, even though the achievement level of most of the students in those teachers’ classrooms was well below minimum standards on an annual basis. Another problem associated with past teacher evaluation systems is the lack of communication between the evaluator and the teacher. In traditional systems, communication is top-down and only the administrator collects evidence of teacher performance during classroom observations, creating an artificial snapshot of a teacher. This hierarchical system is problematic because the teacher is automatically in a passive role and does not contribute to the evaluation. In addition, the very nature of the process leads to a negative climate where the administrator is looking for a weakness in the teacher’s performance (Danielson & McGreal, 2000). If a negative climate exists in a school, teacher behavior in a formal classroom observation will be vastly different because the teacher will feel that it is a high stakes observation that will affect his or her evaluation.
Another factor that affected traditional teacher evaluation systems was the lack of differentiation between novices and experienced practitioners. In most systems, the teacher evaluation process, the standards used and the level of expectation are similar for both first year and veteran teachers. This structure can lead to different attitudes towards evaluation by teachers in different stages of their careers. Peterson (2000, 2004) illustrates that beginning teachers expect and solicit evaluation, because it reinforces what they are doing in the classroom. On the other hand, veteran teachers are much less positive about teacher evaluation because the process and standards used are vague and ambiguous. They perceive their final evaluation rating as depending more on the qualities of the evaluator than on how they teach. As a result, veteran teachers view evaluation as unimportant or irrelevant.

As will be illustrated later in this study, the role of the evaluator is a critical factor in determining the success of an effective teacher evaluation system. Reviews of the research literature of the past 25 years highlight the principal supervisor as playing a strong positive role in successful evaluation systems (Colby, Bradshaw & Joyner, 2002).

In contrast, research has also found that ineffective systems rely heavily on the principal or supervisor as the primary evaluator, creating more of an emphasis on summative, rather than formative evaluations. This is problematic because many teachers are more of an expert regarding their work than the administrators who evaluate them. For example, an administrator with a mathematics background may not have the knowledge base to accurately evaluate a physical education teacher. In a Rand study (Wise, Darling-Hammond, McLaughlin, & Berstein, 1984), researchers found that “the lack of sufficient resolve and competence on the part of principals to evaluate accurately” (p.2) was a major problem in school districts with ineffective evaluation systems. Peterson (2000) stated the following: “Seventy years of research on
principals’ ratings of teachers shows that they do not work well.” Well-designed empirical studies depict principals as inaccurate raters both of individual teacher performance behaviors and of overall teacher merit. These problems go beyond simply developing a better rating form, informing teachers of the items on the rating forms, or improving the training of principals as raters, although these strategies are offered by many who are interested in preserving current practice. (p. 19)

**Effective Teacher Evaluation Systems**

The literature on teacher evaluation systems over the last 40 years has consistently supported two significant findings. The first conclusion is that teachers and administrators have always recognized the importance and necessity for evaluation, even though they have had serious concerns about its implementation and the lack of effect that the results had on teachers, their classrooms, and their students. The second conclusion from the research is that evaluation systems designed to support teacher growth and development using a formative evaluation approach resulted in higher levels of teacher efficacy as well as more thoughtful and reflective practices, while still being able to sustain a high level of accountability and expectations (Danielson & McGreal, 2000). The need for a multiple criteria approach to teacher evaluation is supported by many changing conditions including: (a) changing of roles, responsibilities and relationships among teachers, students, and administrators; (b) increased understanding of how adults learn, grow, and develop; (c) current research on the importance and complexity of teaching; (d) increased focus on differentiating between professional development and teacher evaluation; (e) an understanding of effective professional development programs; and (f) an assessment of traditional forms of supervision. Darling-Hammond, Wise, and Pease (1983) identify five minimal conditions for a successful teacher evaluation system. These conditions
include: (a) teachers and administrators understand the criteria and process for evaluation; (b) teachers and administrators understand how these criteria and processes relate to the basic goals of the organization; (c) teachers perceive that the evaluation procedures are intended to enable and motivate them to improve their performance; (d) principals perceive that procedures enable them to provide instructional leadership; (e) participants understand that the evaluation process allows them to achieve a balance between external controls and autonomy. Moreover, Colby (2001) emphasizes that “the teacher evaluation process needs to be tightly connected to the district priorities and school functions, as well as, provide an evaluation environment that supports ongoing, professional learning.” Pappano (2010) writes in her book *Inside School Turnarounds* that effective teacher evaluation systems must include the examination of criteria beyond student test scores. Frederick Hess (2010) and Pasi Sahlberg (2011) address teacher evaluation from a different perspective. Their focus is on creating an educational culture where teachers are trusted to do their jobs. Sahlberg (2011) in *Finnish Lessons* writes that teachers are held in high regard throughout the Finnish school system. Accordingly, there is a trust factor that minimizes any formal evaluation system, as it exists in the United States. In Finland there is no formal teacher evaluation system. A portfolio of a teacher’s work is the primary evaluation tool used in Finland.

Nonetheless, the RTTT competition has determined that an effective teacher evaluation system should provide multiple criteria including credible teaching standards (“Race To The Top”, 2013). In a standards-based teacher evaluation system, teachers’ performance is evaluated against a set of standards that define a competency model of effective teaching (Heneman, Milanowski, Kimball, & Odden, 2006). Standards serve several purposes in a teacher evaluation: model what teachers are expected to know and be able to do based upon the research
and best practices; describe desirable levels of performance; and clarify the elements of what teachers can expect to improve in over the long term (Kleinhenz, Ingvarson & Chadbourne, 2002). Danielson (2007) describes the importance of having a framework for standards:

A framework for professional practice can be used for a wide range of purposes, from meeting novices’ needs to enhancing veterans’ skills. Because teaching is complex, it is helpful to have a road map through the territory, structured around a shared understanding of teaching. Novice teachers, of necessity, are concerned with day-to-day survival; experienced teachers want to improve their effectiveness and help their colleagues do so as well; accomplished teachers may want to move toward advanced certification and serve as a resource to less-experienced colleagues. (p. 2)

Odden (2004) specifies that a standards-based teacher evaluation system must include a set of teaching standards that describes in detail what teachers need to know and be able to do; a set of procedures for collecting multiple forms of data on a teacher’s performance for each of the standards; and a related set of scoring rubrics that provide guidance to assessors or evaluators on how to score the various pieces of data to various performance levels (p. 127). Peterson (2004) describes multiple data systems to include not only administrator visits, conferences and teacher rating forms, but pupil achievement data, student and parent surveys, peer review of materials, documentation of professional activity, teacher test scores, National Board Certification, action research projects, school improvement participation, and information unique to the individual teacher. Moreover, Heneman and Milanowski (2003) note that a standards-based teacher evaluation system should be a part of a comprehensive human resource management system that includes recruitment, hiring, induction, retaining, promotion and development of teachers.

Furthermore, Danielson and McGreal (2000) emphasize that standards-based teacher evaluation systems should be linked to the mission of the school district, be viewed as a continuous process, emphasize student outcomes, and be supported with adequate resources to be successful.

McGreal (1986) researched 300 teacher evaluation systems and identified nine commonalities
that emerged as best practices in teacher evaluations systems.

One common trait was that positive attitudes and procedures, rather than negative attitudes and procedures, are needed to promote instructional improvement. Another shared characteristic was that evaluation procedures were flexible and that data collection was tailored to the needs of the individual teacher. Another best practice was that teacher evaluation systems were most effective when the focus was on the improvement of specific indicators of classroom instruction and teaching behaviors. The next common attribute was that teachers and administrators identified individualized professional development goals were developed collaboratively. Another common trait was that effective teacher evaluation systems emanated from an agreed upon and clearly articulated definition of teaching that gives the teacher and the administrator a common language. A sixth commonality was that there was a reliance on pre-observation planning, observation of specified behaviors, and a post observation feedback conference. The use of alternative data sources such as self-evaluation, peer evaluation, parent evaluation, student evaluation, student performance, and examination of classroom artifacts was widely used in exemplary teacher evaluation systems. Another commonality was a differentiated system for tenured and non-tenured teachers. Finally, exemplary teacher evaluation systems conducted a complete training program of skills and understandings for all participating teachers and supervisors so that they can effectively implement a growth system (McGreal, 1986).

**Leadership and Implementation**

It is generally acknowledged that teachers have the greatest impact on student achievement (Darling, 1997), while the role that school leaders play is under valued. Primarily, school leaders affect instruction indirectly via the allocation of a reasonable budget, the enforcement of student conduct or the determination of relevant professional development
(Halverson et al, 2004). However, school leaders can also affect teaching and learning through supervision and evaluation. The evaluation process is a formal opportunity to communicate district-wide and school goals, instructional strategies, standards and values to teachers (Wise et al, 1984).

Past teacher evaluation practices were described as inaccurate, lacking in standards and providing insufficient training. In the era of RTTT, standards-based teacher evaluation instruments that use multiple sources of data have been required in all states and districts participating in the initiative (U.S. Department of Education, 2009). Over thirty participating states have adopted the Danielson Model or the Marzano Model framework for teaching and evaluation (National Comprehensive Center for Teacher Quality, 2012).

Despite the availability of research-based evaluation instruments, the key to effective implementation of these models of teacher evaluation lies with the school leaders. Essentially, the implementation of available research-based evaluation instruments lies with school leaders. Leaders can view and implement the teacher evaluation process in two ways: as superficial and meaningless bureaucratic necessity or as a meaningful process (Doherty, 2007).

In the Halverson et al (2004) case study *Implementing teacher evaluation systems: How principals make sense of complex artifacts to shape local instructional practices*, four schools were included in the study with the leaders implementing a standards-based teacher evaluation system based upon the framework for teaching (Danielson, 1994). The case study collected responses from the teachers as well as the school leaders that implemented the evaluations. In summary, the teacher’s responses were generally positive and they were most positive about the feedback they received. Some teachers cited specific examples of the evaluator’s feedback that resulted in changes they made to their practices. School leaders that claimed to follow the
process with fidelity evaluated the teachers that seemed to benefit the greatest from the evaluation process. Most school leaders wanted to make this system work and intended to follow the process with fidelity. However, it was evident that the process took as much as twenty-five percent of their school leader’s time. Halverson et al (2004) stated that they saw the school leaders develop complex schedules and invest much personal time to accomplish the process. Ultimately, the school leaders felt that the considerable time commitment required to complete the evaluations prevented them from implementing the full scope of the process.

Further evidence as a result from teacher feedback from the case study suggests that the evaluators may not have the necessary training to provide worthwhile feedback to high performing teachers. Halverson et al (2004) state: evaluators instead used evaluation as an opportunity to work with novice teachers to build a positive school culture rather than to push instructional practices to the highest level” (p. 36). However, Halverson et al (2004) concluded that it could be discerned from their study that this lack of skill was a cause or an effect of evaluator priorities.

Halverson et al (2004) concluded from this case study that the implementation of the evaluation process varied from school to school and that the process varied by the way the evaluators understood their role and their available capacity.

**Research on Value Added Measures as a Criteria for Teacher Evaluation**

In 2010, the Economic Policy Institute (EPI) published a brief that cited serious concerns regarding the used of student test data alone to evaluate teacher effectiveness. Several noted scholars of education collaboratively authored *Problems with the Use of Student Test Scores to Evaluate Teachers* (Baker et al, 2010). The summation of their research concluded that student test results should be included in the teacher evaluation process, but should only be one piece of
information used to evaluate effectiveness. Essentially, Value Added Modeling (VAM) measures a teacher’s contribution in a school year by comparing his or her students’ test scores within a given year to the same students’ progress in the previous school year along with the results of other students in that same grade. The use of Value Added Assessments is intended to isolate the contribution each teacher makes in that particular school year towards his or her students’ growth.

The Brown Center on Education Policy of the Brookings Institute also sponsored a study on the use of VAM in teacher evaluations. This study, conducted by several well-regarded educational scholars, strongly advocates for the use of VAM in teacher evaluation. While the study recognized the controversy around the use of teacher performance measures that incorporate VAM, it contends that this quantitative data can be useful to teachers (Glazerman, 2010). The study states that teachers receive scant feedback on their past performance in raising test scores and that the information that they do receive pertains strictly to the yearly average test scores. The study also argues that this yearly data can be misleading as opposed to data that measures the progress their students have made from previous years (Glazerman, 2010). While value added evaluations are considered a fairer gauge of teacher effectiveness than tests results from within a given year, the Brown study recognizes that VAM should be used discreetly and with great caution (Glazerman, 2010).

The critics of VAM, as pointed out in the EPI brief, state that the use of VAM to evaluate individual teachers has not yet been scientifically validated. It is also widely recognized that numerous factors can favorably influence student academic growth, aside from a student’s current teacher (Newton, Darling-Hammond, Haertel & Thomas, 2010). In-school factors that can significantly influence academic growth include a student’s previous teacher(s), tutors,
instructional specialists, the quality of curriculum materials, class size, and school attendance (Baker et al, 2010). Out of school learning is also a factor (Baker et al, 2010). Well-educated and supportive parents can help with homework and provide a wide array of enriching educational experiences such as trips to museums and libraries. In contrast, students from economically disadvantaged communities may have parents who are unable to support their academic achievement (Newton et al, 2010). Teachers’ value added evaluations can also be negatively affected by summer loss of learning that occurs more readily in lower economic communities (Newton et al, 2010). Even when statistical methods are used to adjust for the socio-economic division, teachers typically receive poorer VAM evaluations when they teach new English language learners (Baker et al, 2010). Other negative unintended outcomes of value added teacher evaluations include teacher demoralization, lesser teacher collaboration and a disincentive for teachers to work with needy students (Newton et al, 2010). The additional recommended criteria in the EPI study included teacher observations of classroom practices, teacher interviews, and artifacts such as lesson plans, assignments and student work. The report presents the need to employ a comprehensive approach to the evaluation process that provides each teacher with guidance and feedback provided by supportive leadership and in working conditions that can improve performance. Additionally, the process must allow for removing teachers who consistently prove ineffective without imposing a flawed quantifiable system (Newton et al, 2010).

Douglas Harris (2012) concludes in his Carnegie Foundation sponsored study that value added measures are positively related to almost all commonly accepted measures of teacher performance such as principals’ evaluations and classroom observations. Harris (2012) states that it is critical that these value added measures are based upon valid and reliable criteria.
Validity refers to the degree to which something measures what it claims to measure. Reliability refers to the degree to which the measure is consistent when repeated. Furthermore, Harris (2012) contends that value added measures may have less bias than an administrator observation and therefore may be more accurate measurement of teacher effectiveness. However, Harris (2012) cautions that value added measures should be only a part of the mix.

**Research On The Use Of VAM In Untested Subjects**

To date, there is limited empirical research literature available regarding the use of value added measures to determine the effectiveness of teachers of untested subjects. One of the available studies was sponsored by the *National Comprehensive Center for Teacher Quality*. This research and policy brief was developed to help states consider options for assessing student learning growth for the majority of teachers who teach content not assessed through standardized tests (Goe & Holdheide, 2011). This study acknowledges the challenges of using student achievement assessments to measure teacher effectiveness and contends that a students’ prior school experience in the same content area may influence performance.

Based on available research, the authors have developed four options that could potentially determine the effectiveness of teachers in non-tested subjects with only two options being relevant to this research paper. Option three recommends that noncore teachers design their own performance goals in their own area of specialization with their effectiveness based on attainment of these goals (Goe & Holdheide, 2011). The brief cites this option was effectively used in the Denver Public Schools as well as the Orange County School District in Florida. In these settings, teachers collaborate with administrators to set performance goals that are linked to student learning outcomes.
Goe and Holdheide (2011) also recommend a fourth option that suggests the creation of new student assessments as a way to measure teacher effectiveness in noncore subjects. The study cites Florida as a state that has used this approach since 2006. However, the brief concludes that option four may not prove cost and or time effective, as it will require new student assessment instruments that will need to be proved valid and reliable. The study further concludes that option three provides a clear advantage as it encourages teachers to demonstrate excellence in their respective areas of specialization (Goe & Holdheide, 2011). Another advantage of option three is that it avoids the time and expense required to develop, administer and score new student assessment instruments in subjects that are not currently tested. Goe and Holdheide (2011) also cite research that suggests principals / administrators are quite adept at identifying very high and very low performers, but not as accurate at assessing teacher performance within the middle of the range. Their research also suggests that there is a tendency for principals to be overly lenient with their ratings and give it higher ratings than may be deserved (Goe & Holdheide, 2011).

The Tennessee Department of Education in conjunction with the Tennessee Association of Health, Physical Education, Recreation and Dance designed a teacher packet that provides tools to demonstrate student growth in physical education. A committee of experienced current and retired physical educators was charged with designing this document and accompanying assessments. This tool was completed in August 2012 with the intent to be piloted in the 2012-2013 school year. The document was designed to measure student learning over the span of a school year in physical education (Hall et al, 2012). The assessments are drawn from national and state standards. The results of these assessments are intended to provide feedback to physical educators, classroom teachers, administrators, parents and policy makers on the status
of student growth in physical education in Tennessee. The document states that the student growth model is a method of measuring progress towards two points in time. It should be noted that the Tennessee state standards for physical education are aligned with the national physical education standards. The introduction further clarifies that this approach is different from measuring student achievement levels measured strictly against a state or national standard. This approach takes into account where the student started prior to instruction and the resulting growth after instruction has occurred. The committee also stated that this student growth model has the potential to inform instructional improvement, evaluate the effectiveness of physical education programs and target needs for professional development (Hall et al, 2012).

In developing this tool, the committee referenced several documents that are designed to assess student performance using Bloom’s Taxonomy of Learning Domains (Bloom, 1956). These learning domains include psychomotor, cognitive and affective as well as health related fitness. Of the seven documents referenced, only PE Metrics and Fitnessgram were field-tested and as a result are considered valid and reliable assessments. However, the rubric for the PE Metrics tool was adjusted from four levels to five levels and as a result could impact the validity and reliability of the assessments (NASPE, 2010). Fitnessgram is a criterion-referenced assessment of health related fitness and unless used to demonstrate growth as opposed to its intent to meet standards would also present problems with validity.

Conclusions

Based upon the aforementioned research, several conclusions can be drawn about the implementation of value added measures as criteria for determining the effectiveness of physical education teachers and other non-tested subjects teachers. First and foremost, effective evaluation programs use multiple criteria to determine the effectiveness of teachers. The

Secondly, while the use of value added measures is controversial much of the literature argues that it can be more accurate than principal / administrator observations (Ghoe & Holdheide, 2011). Moreover, those researchers who tend to take a more negative view of the use of value added measures as evaluation criteria, can not fully dismiss its’ legitimacy when used in conjunction with additional criteria (Baker et al 2010, Newton et al, 2010).

It was also evident, based upon the research, that the role of the leadership in the implementation of a new evaluation process is significant and can make or break the process and cannot be understated (Halverson, Kelley & Kimball, 2004). Lewin (as cited in Covey, 1990) points out the importance of involving all stakeholders in the procedure of creating change. When administrators involve teachers in the implementation of a new teacher evaluation system, it is more likely that they will be committed to coming up with solutions and likely accelerate effective implementation (Sergiovanni, 2001).

Though the content area of physical education was not a part of the Nation at Risk agenda, nor is it a part of recent education reform efforts such as No Child Left Behind Act of 2001 (U.S. Department of Education, 2002) or even the most recent Race to the Top initiative (U.S. Department of Education, 2010), physical educators have been pulled along with the movements. Despite the marginalization of the profession, the field has arrived at a point in time where upon accountability is expected. This accountability movement has involved all teachers, regardless of the subject area. For the first time in most states, teachers of physical education, health, art and music are involved in the effort.
Regarding the use of value added measures as criteria for evaluating physical education teachers the Tennessee Department of Education in conjunction with the Tennessee Association of Health, Physical Education, Recreation and Dance has designed a teacher packet that has potential to be used as a model for efficiently measuring student growth (Hall et al, 2012). The Tennessee packet includes several assessment strategies that are aligned with their state standards and are capable of tracking student growth over a period of time with recognition for the starting point of each student (Hall et al, 2012). One flaw with the use of this tool is the potential lack of validity and reliability of the assessments drawn from the PE Metrics assessment tool, due to the revision of the rubrics. Harris (2012) stresses the importance of recognizing all assessments as valid and reliable measures of student growth. Nonetheless, by incorporating the true rubrics of PE Metrics, it is feasible that the Tennessee model of measuring student growth in physical education could provide a reasonable prototype for Massachusetts. However, it will be important to study the imminent results from the pilot year of the Tennessee model of measuring student growth in physical education before reaching a conclusion.

**Chapter III: Methodology**

**Research Tradition**

This study is a qualitative bounded case study conducted in a large urban public school district in eastern Massachusetts. The design of this study links the conclusions with the research questions following Lewin’s three-step model of change, Prochaska and DiClemente’s Change Theory. The research questions are:

1. How has the new educator evaluation process that now includes District Determined Measures effected change in the physical educators’ instructional practices as perceived by the physical education teacher?
2. Do physical education teachers perceive any change in their professional relationship with their administrator due to the new evaluation process that now includes District Determined Measures?

A qualitative approach draws conclusions by studying a phenomenon in a natural setting based upon the meaning the individuals bring to the experience (Denzin and Lincoln, 2005). Merriam (2009) states that qualitative research uses words to detail data. This bounded case study investigated the perception of sixteen physical educators relative to the student impact rating included in the new educator evaluation program. The study has included “the voices of the participants, the reflexivity of the researcher, and a complex description and interpretation of the problem” (Creswell, 2007, p. 37). Qualitative research is preferable when the researcher needs to hear the stories and experiences of others to understand the complexity of an issue at a very detailed level (Creswell, 2007).

**Case Study**

Creswell (2007) describes case study research as “the study of an issue explored through one or more cases within a bounded system…through detailed, in-depth data collection involving multiple sources of information…reports a case description and case-based themes” (p. 73). The case study approach is best when the research includes defined cases and seeks to study the phenomenon in depth (Creswell, 2007). Yin (2009) agrees and defines a case study as “an empirical inquiry that investigates a contemporary phenomenon in depth within its real-life context” and “relies on multiple sources of evidence, with data needing to converge in a triangulating fashion” (p.18). Yin (2009) also maintains that case study research is appropriate when the researcher is trying to “describe an intervention” (p. 20). The efficacy of a case study is linked to the quality of the research questions and triangulation of the data (Stake, 2005).
Stake (2005) describes three types of case studies: intrinsic, instrumental, and multiple. The research conducted on District Determined Measures is primarily instrumental as the case is examined “mainly to provide insight into an issue or redraw a generalization” (Stake, 2005, p. 445).

**Sample Design, Participants and Recruitment**

The population that was studied consisted of sixteen K-12 physical education teachers from a large suburban school in eastern Massachusetts. The experience of these physical educators varied from two to thirty years in the district. These physical educators have participated in the new Massachusetts educator evaluation process and have now incorporated the use of District Determined Measures as an indicator of student growth and teacher effectiveness. Moreover, the newly established DDM’s, for physical education, were successfully submitted and approved for use by the Massachusetts Department of Elementary and Secondary Education.

The coordinator / director is responsible for supervision and evaluation of the physical education personnel and possesses fifteen years of administrative experience. The research site had implemented both the summative performance rating and the Student Impact Rating / DDM’s of the new Massachusetts educator evaluation program. Purposeful sampling was used to select participants. To recruit potential candidates the researcher adhered to the following protocols:

- Identified three to four qualifying potential school districts in eastern Massachusetts to participate in the study with anticipation that one district will agree to participate in the study.
• Formal requests to participate were sent to the superintendent of schools of the identified districts via email. The letter outlined the purpose of the study, the procedures, potential risks.

• If a school district was agreeable, then an introductory email was sent to the director of physical education outlining the purpose of the study, the procedures, and the potential risks and inviting the district to participate in the research by scheduling a twenty minute information session via Google schedule tool.

• At the scheduled information session, the researcher further explained to the director the three phases of data collection that will include:
  
  o Document review of newsletters, memos, website and / or announcements
  
  o Observations of meetings on the topic of teacher evaluation
  
  o Semi-structured individual interviews

At this informational meeting, the researcher responded to any questions or concerns presented by the director. Once the district was selected, all K-12 physical education teachers received an introductory email outlining the purpose of the study, the procedures, and the potential risks and invited them to participate in the research by scheduling a twenty to thirty minute information session via Google schedule tool. At the conclusion of this process, an informed consent was provided to all agreeable teacher participants.

Once the district was selected and all participants were assured of confidentiality, a formal request to review relevant documents and a request for a schedule for any impending staff meetings where the topic of teacher evaluation will be discussed was submitted. After the first phase of data collection was completed, the researcher scheduled semi-structured interviews with
all agreeable physical educators and arranged to attend and observe grade level and all staff department meetings that were specific to the topic of teacher evaluation.

Data Collection

Data collection in a qualitative study uses words to detail data (Merriam, 2009). Creswell (2009) states that the most common methods to collect data for qualitative research is through interviews, observations and document review. Yin (2009) further states that there are six methods of data collection for a qualitative case study as follows: document review, archival review, interviews, direct observation, participant observation and physical artifacts (p. 99).

The main source of data for this qualitative study will be based upon individual interviews. Yin (2009) contends that conducting interviews is the best method of gathering data for qualitative case study research. Kahn and Cannell (1957) describe interviewing as “a conversation with a purpose” (p. 149). Marshall (2006) states, “The participant’s perspective on the phenomenon of interest should unfold as the participant views it, not as the researcher views it” (p. 101). Therefore, interviews should be more open ended and less structured (Merriam, 2009). Being a good interviewer requires skill and experience (Thomas, Nelson & Silverman, 2011). Establishing a rapport with the interview respondent is essential to develop a level of trust that enables the researcher to uncover true feelings, thoughts and intentions of the respondent (Thomas et al, 2011). Also, the interviewer must be alert to non-verbal reactions as well as the verbal reaction of the respondents and be flexible in rephrasing while pursuing certain lines of questioning (Thomas et al, 2011). Furthermore, the interviewer must use words that are clear and meaningful to the respondent and be able to ask questions so that the participant understands what is being asked (Marshall, 2006).
However, in an effort to provide an interview that is meaningful it becomes important to gather background information and historical context for the study. The use of documents often entails a specialized approach called content analysis (Marshall, 2006). The content to be analyzed may be any form of written material including pictures, newspaper articles, speeches or emails. Marshall (2006) contends the greatest strength of content analysis is the consideration that the approach is unobtrusive and typically nonreactive. Marshall (2006) considers this aspect of data collection supplemental to other methods and states, “the review of documents is an unobtrusive method, rich in portraying the values and beliefs of participants in the setting” (p.197). As a result, it becomes appropriate for the researcher to review documents that are related to the content; District Determined Measures. The researcher believes that the review of agendas and minutes of faculty meetings, professional development documents, announcements and formal policy statements regarding the implementation of district determined measures may be useful in developing an understanding of the setting and administrative expectations around DDM’s. Marshall (2006) cautions that because the analysis by the researcher requires interpretation it is critical that care is taken in the inferred meaning gathered from these documents.

Finally, participant observation is both an overall approach to inquiry and a data-gathering method that is recognized, to some degree, an essential element of qualitative study (Marshall, 2006). Through this method of data collection, the researcher was truly immersed in the social phenomenon. Marshall (2006) believes the method provides the researcher an opportunity to see, to hear and begin to experience the phenomenon as the participants do. Such an immersion enables the researcher to learn directly from his experience. Glesne (1999) contends that personal reflections are integral to the analysis of a participants perspective and
they provide the researcher new vantage points and with opportunities “to make strange familiar and familiar strange” (p. 112).

**Individual Questions**

Open-ended questions were best suited to collect data that provided the greatest insight into the phenomenon being studied (Merriam, 2009). Questions that best solicited candid responses opened with statements such as: a) Tell me about when….b) Give me an example of…..c) Tell me more about that….d) What was it like for you when…..(Merriam, 2009, p. 99).

The questions asked for this study followed closely Merriam’s recommended strategies when conducting a semi-structured format for individual interviews. Furthermore, Merriam (2009) recommends that interviewers avoid multiple – part questions, questions that evoke assumptions or questions that can be answered with “yes” or “no” responses. All individual interviews were recorded with an iPhone 5s using an inexpensive yet highly effective and user friendly commercial application that was specifically designed for recording research. The commercial application allowed all interviews to be sent electronically to a professional transcription provider affiliated with the application.

Research questions were developed to articulate what the researcher wants to know and understand. Creswell (2005, p. 92) states that actual interview questions are what the researcher will ask the participants in order to gain their understanding of the phenomenon (Creswell, 2005, p. 92). A productive interviewer must be able to ask relevant questions and have the ability to interpret responses while being an active listener (Yin, 2009). The researcher utilized a semi-structured interview process. Merriam (2009) states that semi-structured interviews mix structured pre-determined questions with less structured questions. In essence, the interview process was conducted similar to a conversation with open-ended and flexible questioning.
Individual interviews were conducted when the participant was available: before school, off periods or after school. The interviews were conducted in person and a pseudonym was used to describe each participant in this study.

**Data Analysis**

The focus of the data analysis was intended to gain an understanding of the social phenomenon surrounding the use of DDM’s from an insider’s perspective relative to two main issues: a) the perceptions of physical education teachers and administrators around the use of DDM’s as a catalyst for change in teaching behavior, b) the perceived change in the professional relationship with the supervisor due to a new teacher evaluation program.

Seidman (2005) believes that the conventional way of presenting and analyzing data is to organize it into categories or themes. He also cautions researchers to avoid analyzing the data until all interviews are completed. Seidman (2005) also believes that if the researcher is conducting the interviews him/herself that they proceed with an open mind and allow the data to emerge spontaneously. Seidman states further, “At the same time, no interviewer can enter into the study of an interview with a clean slate” (Seidmen, 2005. p.117). Accordingly, it is incumbent upon the researcher to identify his or her interest in the subject and make sure that the interest is not infused with bias, anger or prejudice (Seidman, 2005). Therefore, the researcher must view the transcript without a predisposed attitude and allow the interview speak for it self (Rowan, 1981 as cited in Seidman, 2005).

After each interview, the researcher forwarded the recording to the transcription provider. Once all interviews were transcribed, the researcher forwarded each participant a copy of the transcription for approval of accuracy. Also, the researcher closely reviewed all field notes accumulated from the observations and made every attempt to insure no bias or prejudice was
present by submitting them to the supervisor. Once approved, the researcher began the process of developing themes through the use of “emotion coding”. This method of coding is described in Miles, Huberman and Saldana (2014) as a method that labels the emotions described by the participant or inferred by the interviewer and will be considered as stage one in coding process. Yin (2014) suggests that the researcher read and re-read the transcripts multiple times prior to placing data into themes or codes.

The stage one coding is followed by “memoing” also described by Miles et al (2014). An illustration will be developed for this process and an example is shown in Table 1 below. This method of coding would be considered to be employing an “inductive approach” in order to develop a *casual network* or *casual chain*. In this case, the interviewer will be inferring the emotions of the participant.

*Casual Chain: Table 1*

<table>
<thead>
<tr>
<th>Change initiative</th>
<th>Recognition of Need, Concern</th>
<th>Resentment to process</th>
<th>Disappointed, Not enthusiastic</th>
<th>Enthusiastic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change in student assessment practice must occur to meet new educator evaluation process: DDM’s</td>
<td>Participant recognizes the need, but concerned about the work ahead to implement the process.</td>
<td>Participant is required to participate in professional development workshops in order to develop DDM’s only to be revised by administration.</td>
<td>Participant disappointed in the process of selecting the DDM’s and the level of input afforded the teachers implementing the change.</td>
<td>Participant enthusiastic about the potential to impact on the ability to demonstrate student learning, yet still resentful over the process.</td>
</tr>
</tbody>
</table>

**Summary**

As previously cited, data collection occurred in three phases as follows: document review, observations and semi-structured interviews. Table 2 provides the process of analyzing the data.
**Data Analysis: Table 2**

<table>
<thead>
<tr>
<th>Task</th>
<th>Process</th>
<th>Product</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data collection Phase 1, Document Review</td>
<td>Review relevant public documents</td>
<td>Notes pertaining to administrative perspective on new evaluation program</td>
</tr>
<tr>
<td>Data collection Phase 2 Observation</td>
<td>Attend meetings w/ staff that relate to new educator evaluation program</td>
<td>Field notes on discussions, questions and body language</td>
</tr>
<tr>
<td>Data collection Phase 3 Interviews</td>
<td>Conduct semi-structured individual and focus group interviews.</td>
<td>Audio recording of responses to primary, secondary questions and unstructured conversations</td>
</tr>
<tr>
<td>Data Analysis 1</td>
<td>• Transcribe semi-structured interviews</td>
<td>Data ready for coding</td>
</tr>
<tr>
<td></td>
<td>• Transcribe field notes from observations and document review</td>
<td></td>
</tr>
<tr>
<td>Data analysis 2(a) of physical education teachers and administrator responses</td>
<td>Code data utilizing inductive approach based on Miles et al (2014) “emotion coding”</td>
<td>Data coded is aligned and prepared for cross coding with change theories as feasible</td>
</tr>
<tr>
<td>Data analysis 2(b) of physical education staff responses</td>
<td>Code data utilizing deductive approach based upon six stages of change as pertaining to Prochaska and DiClemente’s Change Theory</td>
<td>Data coded is aligned and prepared for cross coding with change theories as feasible</td>
</tr>
<tr>
<td></td>
<td>• Pre-contemplation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Contemplation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Planning</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Action</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Maintenance</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Termination</td>
<td></td>
</tr>
</tbody>
</table>
| Data analysis 2(c) of physical education administrator response | Code data utilizing deductive approach based upon the three domains of Lewin’s three-step model of change | Data coded into deductive themes:  
- Step one – unfreeze  
- Step two – movement  
- Step three – refreezing |

| Data Analysis 3 | Data interpreted | Results – sense making and interpretation of responses |

**Trustworthiness**

Merriam (2009) points out that case studies are limited by the integrity of the investigator and the commitment to use all the data that was collected as part of the study. In the past, case study research has been criticized for lack of rigor, investigator bias, and difficulty in generalizing a case to the larger population (Merriam, 2009). She suggests, “The inclusion of multiple cases is, in fact, a common strategy for enhancing the external validity and generalizability of your findings” (p. 50). The ethical conduct of a researcher conducting a qualitative research project will reduce any threats to validity and reliability (Merriam, 2009). Maxwell (2005) states, “Validity is a goal rather than a product; it is never something that can be proven or taken for granted” (p. 105). Maxwell (2005) argues that the collection of solid evidence will help to mitigate possible concerns about validity.

Mears (2009) cautions the researcher about objectivity and bias in conducting interviews. There is the potential for the researcher to lead the participant in certain directions with both questions and follow-up inquiries. Having a background in the subject matter can either increase or diminish the likelihood of bias; accordingly, caution must be taken when conducting the research in a site familiar to the researcher, such as the workplace (Mears, 2009). “Research from within the setting becomes more challenging, for it requires overcoming your personal lens in order to understand from other’s point of view” (Mears, 2009, p. 83).
The validity and reliability of the study results is a significant responsibility of the researcher (Creswell, 2007). The researcher recognized that internal biases had potential to be present within the study and affect the chosen school district. As a result, it was important to recognize potential influences and to use it productively (Maxwell, 2005). External validity explores how the results from one study align with the results from other similar studies within a wider population of administrators and teachers outside of the locale of this study, eastern Massachusetts. Reliability infers that any other researcher undertaking a very similar study will follow similar procedures followed. To ensure that the study is valid and reliable, the data will be triangulated.

**Data Storage**

All data has resided with the researcher and was stored electronically on the researcher’s computer. It was backed up daily to an 8 GB external storage device and then backed up to an external hard drive twice a week. Once all participants were provided the opportunity to review their transcripts, they were destroyed.

**Protection of Human Subjects**

The National Commission for the Protection of Human Subjects of Biomedical and Behavior Research (1979) specifies the ethical principles for the protection of human subjects, which are (a) respect for persons; (b) beneficence; and (c) justice. Respect for persons includes the right to volunteer to be part of the study, to be fully informed, and to protect those who have diminished capacity. Beneficence requires the participant to be protected from harm and to ensure that all participants are cared for at all times. Justice refers to ensuring that all participants are selected as well as treated equally, ethically, and fairly.
In order to participate in this study, all participants completed an informed consent document. Additionally, each participant was given a pseudonym to protect confidentiality. Participants were informed that all data will be held in strict confidence, and that all audio recordings will be destroyed upon conclusion of the research.

**Chapter IV: Report of the Findings**

This chapter presents and discusses the key findings of this research study conducted during the late winter and early spring of 2015. The chapter includes a review the study context and provides an overview of the site and the research participants. The following sections will present the themes based upon the review of pertinent documents, observations, and individual interviews as related to the following research questions:

1. How has the new educator evaluation process that now includes District Determined Measures effected change in the physical educators’ instructional practices as perceived by the physical education teacher?
2. Do physical education teachers perceive any change in their professional relationship with their administrator due to the new evaluation process that now includes District Determined Measures?

The final section of this chapter presents a summary of the most salient research findings.

**Study Context**

This study explored the perceptions of physical educators regarding the Student Impact Rating, which is based on District Determined Measures of student performance. The researcher believed that it was important to explore the physical educators’ beliefs, feelings, and perceptions regarding this aspect of the evaluation process due to the necessary assessment practices that are not often employed in physical education.
Research site

The superintendent of schools for the participating district granted approval in January of 2015 for the research site of preference. The site was preferred as it met all of the established parameters. The research site offered a staff of at least 20 physical education teachers that possessed a wide range of experience as practitioners. The site also employed a content trained supervisor that evaluated the staff. Furthermore, the district physical education program had developed and implemented two District Determined Measures. The selected DDM’s were submitted and approved for use by the Massachusetts Department of Elementary and Secondary Education. The school district has a racially and economically diverse population of about 63,000. There are 5,264 students in grades pre K – 12. The student population is 41% white and the remaining 59% are Hispanic, Asian or African-American. Based on information secured from the Massachusetts Department of Elementary and Secondary Education website, 41% of the student population is considered low income (free or reduced lunch). Seventeen percent of the entire student population is comprised of English language learners. The student population has continued to grow over the past five years. The population increase has primarily consisted of English language learners and homeless students.

There are nine schools in the district, with six elementary, two middle schools and one comprehensive high school. All but the high school have been built or renovated within the last fifteen years. There are plans underway to renovate or build a new high school. The teaching facilities for physical education are excellent. Each elementary school has a full size gymnasium that can be divided with a curtain if needed. All six elementary schools have a horizontal climbing wall. Each elementary school has ample outside field space as well. One middle school has a fieldhouse like gymnasium with three teaching stations divided by curtains as
needed. However, it has had limited outside field space. The other middle school has a large gymnasium with ample outside area for activity. Both middle schools have vertical climbing walls with several high rope elements used for their adventure education units. The high school has several teaching stations. They include two large gymnasiums, a weight / fitness room with abundant free weights, and cardio equipment. The high school also has a multi-purpose room that can be used for a wide range of physical education experiences that could include floor based fitness experiences such as yoga, zumba, step aerobics, Pilates and dance. There is an inside climbing wall as well as an extremely well designed high and low ropes outdoor challenge course. It is considered one of the most unique in New England. The middle schools and the high school also boast outstanding technology that consists of an apple TV set up at each school, LCD projectors at each school and approximately thirty H-7 heart rate monitors per teacher. The technology is used regularly to encourage students to have an increased level of moderate to vigorous level of physical activity during class time.

In 2011, the school district was awarded a federal grant, the Carol M. White Pep Grant, for more than 1.4 million dollars over three years to support improved physical education experiences that encourage an active lifestyle and healthy nutritional behaviors. The Carol M. White Pep Grant also allowed the program to revise their physical education curriculum and provide students an articulated and sequential K-12 physical education experience. Assessing student performance was also an initiative of the grant that aligned well with the implementation of DDM’s.

As previously stated, the physical education staff at the research site consisted of twenty physical educators. Purposeful sampling was the process the researcher utilized to recruit participants. Once the site was approved, an email was sent to all physical education staff within
the district explaining the study. The email also solicited interest among the physical education staff to contribute towards this study by agreeing to meet for a one on one interview at their convenience. Potential participants were assured of confidentiality.

Sixteen of the twenty available physical educators agreed to participate in the study. The participant pool was diverse and consisted of seven males and nine females. The length of service as physical educators was also diverse and ranged from four years to thirty years of service as physical educators in Massachusetts. Twelve participants had received their Masters Degree in physical education. All participants were fully certified and possessed professional licensure. All but one participant had attained professional status in their district.

There are six elementary schools in the district. The elementary physical education staff consisted of seven teachers. All but one has an advanced degree. The entire K-5 staff agreed to participate in the study. There are four females and three males within this group. Elementary physical educator, Susan, has taught for thirty years in the district. Larry is another veteran teacher and has accumulated twenty years in the profession with fifteen in the district. Jackie has taught for twenty years in the district. Joe has seventeen years of service with fifteen years teaching at the elementary level. Lisa possessed eleven years of experience with all of them in the district and at the elementary level. Mary has been teaching for six years. She taught one year at the middle school and the remaining at the elementary level. David has taught for a total of 4 years, but only two in the district and at the elementary level.

There are six middle school physical education teachers in the district servicing two middle schools. Only five of the six physical educators were eligible to participate in the study based on proper certification and one middle school teacher opted not to participate. Of the remaining four teachers two from each school agreed to participate. The middle school
participants were evenly divided by gender. Three have advanced degrees. The most veteran middle school teacher is Marcia and she has taught for thirty years in the district at varied levels. Will has taught for nine years in the district and all at the middle school level. He also coaches three varsity sports in the district. Jim is a nine-year veteran of the district. Liz has taught physical education in the district for eight years. She also coaches at the high school.

The high school program has a large staff consisting of seven teachers. There are four male teachers and three females. All three females opted into the study and two of the males agreed to participate in the research. Four have advanced degrees. Danielle has taught for twenty-three three years in the district and coaches varsity sports at the high school. Rob has taught for twenty years, all within the district, and also coaches varsity sports. Barbie has taught sixteen years in the district. Lee has taught physical education for eleven years and seven of those in the district. Leo has taught six years at the high school level and coaches at the high school.

The curriculum director / supervisor, Fred, is also located at the high school. This past school year was Fred’s thirty-fifth in the district. Prior to his administrative role, Fred taught physical education for twenty years at each grade level within the district. He was also recognized as an extremely effective physical educator. As a result, he has earned credibility from his teaching staff, as a supervisor. It is important to note that his position is multi-faceted, as he is also the director of athletics. This dual responsibility is not unusual in public schools in Massachusetts. However, this additional responsibility requires a great amount of his professional time. Most curriculum directors that also oversee athletics have the benefit of a full time secretary. Unfortunately, that is not the case in this district as Fred shares his secretary with at least five other curriculum directors / department heads at the high school. As a result, Fred
must balance numerous responsibilities, yet has the reputation, among his peers, as being a leader in the content area of physical education as well as an excellent director of athletics.

The research site initiated the implementation of DDM’s in the fall of 2013. The district followed the initial implementation plan as described by the Massachusetts Department of Elementary and Secondary Education timeline. Table 3 provides the original and the revised timelines with an extension possibility for implementing the Student Impact Rating in the new educator evaluation program. It has been reported that many districts have yet to have the untested content areas pilot DDM’s (personal communication, MA DESE, November 16, 2014).

*MA DESE original and revised DDM implementation timeline: Table 3*

<table>
<thead>
<tr>
<th>Original Timeline</th>
<th>2013-14</th>
<th>2014-14</th>
<th>2015-16</th>
<th>2016-17</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collect data</td>
<td>Collect data</td>
<td>Collect data</td>
<td>Collect data</td>
<td>Collect data,</td>
</tr>
<tr>
<td>(Year 1)</td>
<td>(Year 1)</td>
<td>(Year 2),</td>
<td>(Year 2),</td>
<td>Issue Student</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Issue 1st</td>
<td>Issue 1st</td>
<td>Impact Rating</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Student Impact Ratings</td>
<td>Student Impact Rating</td>
<td></td>
</tr>
<tr>
<td>Revised Timeline (as of April 2013)</td>
<td>Research and pilot potential DDM’s</td>
<td>Collect data (Year 1)</td>
<td>Collect data (Year 2), Issue 1st Student Impact Rating</td>
<td>Collect data, Issue Student Impact Rating</td>
</tr>
<tr>
<td>Extension Request</td>
<td>Research and pilot potential DDM’s</td>
<td>Collect data (Year 1)</td>
<td>Collect data (Year 2), Issue 1st Student Impact Rating</td>
<td></td>
</tr>
</tbody>
</table>

The use of performance based student assessments that are required of DDM’s, are not typically incorporated in physical education programming. Therefore, adjustments to teacher practices were expected in order to comply with this aspect of the educator evaluation program.

Based upon DESE information, the district-wide supervisor of physical education introduced the concept of District Determined Measures to all Pre K-12 physical education staff in October of 2013. The physical education teachers met monthly by grade levels that included Pre K – 5, grades 6 – 8 and grades 9 – 12. Additionally, quarterly departmental professional
development release times (PDRT) allowed for Pre K-12 staff to meet collectively. The review of agendas for staff meetings indicated that throughout the 2013-14 school year (year one) and 2014-15 school year (year two), DDM’s were listed as one agenda item, if not the entire meeting focus, for each meeting.

The initial introduction to staff provided an overview of DDM’s. The supervisor explained that each grade level was charged with designing two grade level specific assessments to be used as DDM’s. In month two of the implementation, the K-12 physical education staff met twice, once by grade level and as a K-12 staff at a PDRT meeting. The supervisor scheduled several trainings for the PDRT meetings. A Physical Education Teacher Education program of a local university presented the trainings. At the meetings, the K-12 staff was introduced to PE Metrics, a standards based cognitive and motor skill assessment package that uses valid and reliable evaluation tools to measure student progress towards achieving the national standards for physical education. The PE Metrics package included video demonstrations, rubrics and directions for providing a setting that is conducive to conducting reliable assessments of student cognitive and motor skill performance. Sponsored by SHAPE America, the national organization for health and physical education professionals, this tool strongly recommended use of video to record and then review student performance of motor skills to align with the accompanying rubric. At this training, each physical educator was provided an iPad mini to record student performance with instruction on proper use of the video app. At the grade level meetings, each staff member agreed to select an assessment either from the PE Metrics package or a self-designed assessment. If a physical educator chose a PE Metrics assessment they reviewed the associated rubric and implementation process. If the assessment was self-designed, the staff member was charged with developing a rubric and implementation document. The self-
designed assessment was required to be approved by the assistant superintendent of schools. Each physical educator was expected to pilot his or her specific grade level assessment and report back in April of 2014. In May of 2014, each grade level selected two DDM’s based upon a group discussion, of those piloted. The selection was determined by three guiding questions: a) how challenging was it to implement the assessment, b) was the assessment developmentally appropriate for recommended grade level, c) was there potential for demonstrating student growth.

The final phase of the DDM implementation was the process used to determine the degree of student growth. During the second half of the 2013-14 school year, the supervisor provided professional development that was intended to clarify the determination of low, moderate or high student growth that was consistent with all grade levels. The guiding document for this procedure was the “Tennessee Model of Physical Education Student Growth Measures” (PE-SGM, 2012). This document was developed and piloted through a collaborative effort between the Tennessee Association of Health, Physical Education, Recreation and Dance and the Tennessee State Department of Education. This model incorporated five levels of performance that included 1) significantly below level, 2) below expectations, 3) at expectations, 4) above expectations, 5) significantly above expectations. The Tennessee PE_SGM model also provided rubrics aligned with specific motor skills and cognitive assessments that were derived from PE Metrics. However, PE Metrics only allowed for four performance levels. The Tennessee Model required a collection of pre and post evidence from an entire class. The Tennessee Model made an assumption that students should grow at least one level per school year in the pre-determined assessments. The final determination of growth was based upon the difference between a
students’ pre and post assessments. Finally, the difference of growth of each student was totaled and then averaged in order to determine the actual measurement of student growth.

At the final 2013-14 school year PDRT meeting, the department supervisor led a discussion with the K-12 physical education staff about the process to determine student growth. At this meeting, all staff members unanimously agreed to utilize the process to measure student growth as described in the “Tennessee Model of Physical Education Student Growth.” This agreement required that all assessments used as DDM’s be revised to include five performance levels. Subsequent grade level department meetings were dedicated to discussing those assessments that were selected as DDM’s that would eventually be reported to the Massachusetts Department of Elementary and Secondary Education. At subsequent grade level meetings, each staff member reported out on their experience with implementing the two selected DDM’s, using the three criteria that determined the common measures. The results of the DDM’s for the 2014-15 school year will be submitted to the Massachusetts Department of Elementary and Secondary Education for state wide analysis.

Findings

In order to understand the perceptions of physical educators regarding the incorporation of DDM’s as a method to determine teacher effectiveness by measuring student growth, the researcher reviewed pertinent documents, observed faculty meetings and conducted one on one semi-structured interviews with sixteen physical educators. For the purpose of this study, the researcher categorized teachers as recent graduates if they had completed their undergraduate or post-graduate coursework ten or less years ago. Those participants that had completed all coursework beyond ten years were described as veteran teachers. The following findings are
primarily based upon the one on one semi-structured research questions and associated sub-questions.

**Research Question 1**

How has the new educator evaluation process that now includes District-Determined Measures effected change in the physical educators’ instructional practices as perceived by the teacher and by the administrator?

Three major themes emerged from the data collected through research question one, and the associated sub-questions. Interviewees contributed differing amounts of information to the three themes that comprise the narrative. Some participants’ spoke at length on one or two themes; some participants contributed to all three themes. As a result, all of the participants’ voices and views are represented in this study. Table 4 presents these themes.

Table 4  
*Themes in Response to Research Question 1.*

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Anxiety has increased as a result of the DDM Requirement</td>
</tr>
<tr>
<td>2.</td>
<td>Preparation has been impacted</td>
</tr>
<tr>
<td></td>
<td>a. University preparation varied</td>
</tr>
<tr>
<td></td>
<td>b. Professional development adequate</td>
</tr>
<tr>
<td></td>
<td>c. Daily preparation around lesson planning</td>
</tr>
<tr>
<td>3.</td>
<td>Need for change not initially recognized</td>
</tr>
</tbody>
</table>

**Anxiety Has Increased as a Result of the DDM Requirement.** During the one on one interview process the theme of anxiety was prevalent throughout, yet to varying degrees.

Feelings of anxiety were explained by the participants in several ways and described in varied
degrees of intensity. A majority of the interviewees that indicated they were anxious about the use of DDM’s used the word “pressure”.

Other terms around the theme anxiety included “nervous”, “stress”, “pressure” and “scared”. In general, most participants did not confide that a high degree of anxiety existed. Yet, most participants did convey that at one point or another throughout the DDM implementation process, they were anxious over the expectation that students must show growth. In particular, the actual implementation of the DDM’s as well as the determination of low, moderate and high growth, were each considered stressors to the process by the participants.

Those physical educators that expressed increased pressure in their practice did so by intimating that the demands of the new educator evaluation program required more thought and had taken time away from their regular lesson preparation. In some instances, participants stated that the use of DDM’s along with the new educator evaluation program has doubled has greatly increased their levels of anxiety. Elementary physical educator Susan, a long time veteran in the study site stated: “When I first heard about DDM’s I was scared….not so much now that I am comfortable with the process.” Contrasting opinions also existed regarding a change in teaching practice, yet pressure still existed. High school teacher, Russ commented:

It (DDM’s) hasn’t changed the way I teach to be honest with you. It seems like it has put on a lot more pressure. Not just on myself, but I think throughout the entire building. It’s just a lot more stress; there have been so many things thrown at us in the last five years. It hasn’t made me a better teacher.

When the researcher asked those participants that had expressed angst what had precipitated their anxiety, eleven of sixteen participants responded that changing the evaluation process to include evidence that their students must have shown growth was the main catalyst. They also implied that they did not necessarily see the need for such a requirement. Elementary physical educator Larry, another veteran, felt it “wasn’t necessary” to formally track his
students’ growth. He stated further: “I don’t feel that I really have to show it…it happens anyways.”

Additional concerns that triggered anxiety among the participants tended to be specific by grade level. The elementary physical educators were concerned that minimal contact time with their students would prove detrimental to demonstrating student growth. Elementary physical education specialist David states: “I feel it (student growth) could be tough just because we only have them (students) once a week for fifty-five minutes. If they don’t practice the skill at home or at recess they aren’t likely to improve.” Another elementary physical educator, Lisa expressed a compelling perspective as follows:

…you have them once a week. If they were in a math class, are you going to teach them the lesson on Friday and then have them come in the next Friday and expect them to remember what they’ve learned and be able to build upon that? To me that’s a big problem.

Secondary level physical educator Barbara, recent recipient of her Masters in physical education, believes that the process has been very stressful, especially at the start up. She felt that there was too much, too soon expected from the administrators and teachers. This resulted in frustration that caused a great deal of stress, and she thought “if you didn’t get it right you could lose your job.”

At a secondary level department meeting in year one of the DDM implementation, secondary level physical education teachers expressed anxiety over a DDM that they had developed based upon the new curriculum model, Adventure Education. They expressed concern that because the DDM was expected to measure improved communication skills, it would be a challenge with an influx of English language learners into the district. They felt it would be difficult to show growth because of the language barrier that existed with ELL students
and therefore reflect poorly on student growth. This was a concern that their supervisor had stated would be researched and reported back to them as soon as possible.

**Preparation Has Been Impacted.** The participants of this study frequently brought up the theme of preparation. Because the connotation of this theme was varied, it was broken into three sub-themes that include: a) university preparation, b) district based preparation via professional development and c) preparation around lesson and unit planning.

The most frequently discussed sub-theme was preparation as it applied to the experiences that the participants had in their physical education teacher education program, whether in undergraduate or graduate level coursework. A sub-question that explored the participants’ comfort level with the use of performance-based assessments, used with District Determined Measures, was the catalyst for this sub-theme. Of the sixteen physical education teachers that took part in this study, only the very most recently trained physical educators could unequivocally state that they felt prepared to implement performance based assessments.

**University Preparation Varied.** Middle school teacher, Liz, who completed her undergraduate coursework eight years ago, expressed much confidence. She felt “100% prepared” and was fine with incorporating performance-based assessments. She stated further “I thought that my undergraduate work was awesome for the field that I am in. I think my undergraduate coursework was very up-to-date…. when I left undergrad, I felt that I could teach that day”. When asked about her grading practice, secondary physical educator, Liz, explained that she incorporated a five point daily scheme that is based on effort, attire, activity level, and attitude. Of the four criteria, only activity level would be considered performance-based. Of course, the use of a pedometer, accelerometer or heart rate monitor would be necessary to provide evidence, which Liz and her colleagues do employ periodically. She stated further that
each of the aforementioned grading criteria is weighted equally. As a result, compliance based grading appeared most prominent in the scheme, as opposed to the performance-based model necessary for use in DDM’s.

Leo, a high school physical education teacher and Mary, an elementary physical educator, are more recent graduates and attended the same undergraduate program. They both felt very confident in applying performance-based assessments. Yet, when asked if they incorporated performance based assessments prior to implementing DDM’s, both teachers stated that they have typically incorporated more subjective and compliance based grading practices based upon effort displayed and wearing proper attire. When asked why they chose to use subjective and compliance based grading schemes, both inferred that they merely followed past practices that were in place when they started their positions. David, the most recent graduate, attended the same university as Leo and Mary. He also felt well prepared with respect to incorporating assessments that are performance-based. He believed that he was “absolutely” prepared in his undergraduate coursework. When asked about his current grading system David stated: “we grade them on psychomotor (domain), how they treat their classmates and if they follow the rules which shows responsibility.” Once again, the majority of David’s grading criteria are classified as subjective, unless a rubric is employed. He confided that he is also following the grading practices that were present when he started as a physical educator in the district.

Veteran physical educators were more apt to have a perception around the “preparation” theme that contrasted with the younger staff members. In general, this category of physical education teachers overwhelmingly believed that they were vastly underprepared to incorporate assessment practices that are required for implementing DDM’s with fidelity. When asked how
well prepared he felt for implementing the assessment process necessary for DDM’s, Larry contended there was absolutely no preparation for this type of assessment in the 1970’s. Larry’s commentary is representative of his veteran colleagues. He stated: “When I went through college, all we did was learn about grades like A, B, C, D. We had no real experience with assessing. It was all about grading students on effort and wearing the right gym clothes.

**Quality of professional development.** The second sub-theme around preparation was directed towards the professional development provided by the district around the implementation of DDM’s. All of the sixteen participants indicated that they felt that the professional development was adequate. All but three participants felt the process was rushed and there could have been a lot more clarity. The three participants that claimed there were no problems with the professional development taught at the elementary level. The K-12 staff that was most vocal about the quality of the professional development, taught at the secondary level. Veteran high school teacher, Danielle directed the following comments towards the Massachusetts Department of Elementary and Secondary Education:

I don’t think they were very clear from the beginning. Their expectations were unclear. I don’t think they actually had physical education teachers involved in the planning of DDM’s. In fact, I understand that the think tank that they (DESE) hired for physical education DDM’s examples actually asked for athletic coaches to be involved. Go figure. We’ve pretty much had to figure everything out for ourselves.

Lee, a secondary teacher, concluded that the physical education department may have rushed the process. She confided, “We were frustrated.” Lee continued her sentiments as follows:

Our department was introduced to it (DDM”)s) even before English or Math, and they were like ‘what are you talking about?’ We were like ‘we have to do DDM’s’ and they’re like ‘what are DDM’s?’ …. there weren’t always answers.
High School teacher, Barbie agrees with her colleagues’ contentions that the DDM rollout could have been done better by DESE. She commented:

I just wish whoever is rolling them out, in the Department of Ed or whoever else, has ironed out the details and knows them inside out…we are told do it this way and then two weeks later we were told something different.

Larry offered a contrasting perception on district-based preparation: “Oh I think I am very well prepared. I don’t always get it right off. But after a few meetings I was fine.” Elementary physical educator, Susan, shared her perception and stated: “I felt pretty prepared…. we were a little bit ahead of the curve and very prepared.” Liz, who had previously cited strong undergraduate preparation, contended later in her interview:

We don’t have assessments like math or English could look at for DDM’s …I think that if we really understood what we were doing we could’ve figured it out on our own. I just think it was new…. I don’t think everybody understood it.

Daily preparation around lesson planning. The third sub-theme on preparation was focused on lesson planning. Eleven of the K-12 staff reported that they were more likely to develop lesson plans that provide an opportunity for their students to practice the District Determined Measures that were established for their grade level. This was especially true at the elementary level. All seven of the K-5 physical educators contended that after conducting the pre-assessment they would regularly plan and provide opportunities for their students to practice the DDM until the post-assessment occurred. They stated that this practice was often at the expense of other experiences that they typically included prior to the requirement to measure student growth. Two elementary physical educators, Larry and Lee, referred to this practice as “teaching to the test.”
Need For Change Not Initially Recognized

Throughout the semi-structured one-on-one interviews it was difficult to distinguish the participants perceptions of change as it related strictly to the implementation of District Determined Measures. While the researcher attempted to maintain the interview focus of change specific to DDM’s, the participants’ perceptions would often drift into the Summative Performance Rating element of the new educator evaluation program that included mini-observations.

When participants were questioned about change that was specific to their practice, as a result of DDM’s, ten participants cited specific changes to their practice. The remaining six participants responded that there was no significant change in their teaching practice when first questioned. Nonetheless, those six particular participants did eventually cite specific changes to their practice in subsequent questions later on in the one-on-one interviews.

Leo, a recent graduate, cited no change. However, later on in the interview he elaborated about the attention that he now places on the mastery objective of a lesson. John explained that he now ensures a formative check-in for understanding throughout the class and especially at the end of each lesson. Another veteran teacher, Susan, contended that there was no change in her practice. Later in the interview she readily explained that she is now more apt to ensure that her lesson plan includes an opportunity for her students to practice the DDM specific to her grade level. Elementary physical educator, Jackie is a veteran teacher that feels there has been no change to her teaching practice as a result of the new educator evaluation program. She believes that an unintended outcome of the new process has isolated her from the building principal. She points to the increased responsibilities that administrators have added due to the new evaluation program. Upon follow-up questioning, Jackie cites collaboration with her colleagues as a major
change in her teaching practice. She specifically cites the staff meetings where they were tasked with collaborating on the selection of DDM’s. Jackie also cited those meetings when the staff collaborated on how to quantify the measurement of student growth. She feels that this collaborative effort has expanded to other areas of professional growth unrelated to the new evaluation system.

The staff that cited a change in their practice readily alluded to a higher level of accountability. As a result, they were admittedly more likely to now “teach to the test,” a behavior unique to physical educators. Larry said: “I purposely now teach to the DDM. It is an activity almost everyday, because I want the kids to improve, so it makes me look good.”

Lisa, an elementary physical educator, believes that she incorporates significantly more assessment in her daily practice than before DDM’s were a requirement of the new educator evaluation program. Lisa especially cited the use of the national physical education standards as a criterion. She was also more likely to video her students as documentation for grading purposes. Lisa was the only physical educator to mention the use of video for documentation. Lee, a veteran, claims that the new educator evaluation plan keeps her on top of things and finds that she is more likely to seek out new resources. She also believes that she is more likely to collaborate with her colleagues due to the new educator evaluation program. Lee specifically cited the process of developing and piloting DDM’s. She noted an enhanced level of cooperation with her colleagues that is necessary to determine which assessment should be chosen for year two of the implementation.

Barbie believes that the data collected for determining student growth has been useful and used the information to improve her teaching. She intimated that when piloting one of the new DDM’s, it was surprising to note that there was no significant growth from the pre-
assessments to the post assessments. She felt that by analyzing this data she was able to make positive adjustments to her teaching practice. Ultimately, she determined that checking for learning throughout the lesson and unit is critical, and has since included these regularly in her practice. Larry feels “the DDM’s have pushed me” and believes that the new evaluation process has made him a better teacher. He claims that he has included new teaching strategies such as posting his agenda and checking for understanding. Joe, a veteran elementary physical educator, says “I’m fine with it” and believes that DDM’s have changed teacher accountability and that it is a reasonable expectation. Another comment about the change in accountability came from elementary physical educator Susan. She believes “it is a change for the better.” Susan thinks that the requirement for tracking student growth holds her accountable to her students.

**Research Question Two**

Do physical education teachers and administrators perceive any change in their professional relationship due to the new evaluation process that now includes District Determined Measures?

Once again, the interviewees contributed differing amounts of information to the three themes that comprise the narrative. All participants’ spoke on at least two themes. Some participants spoke at length, while others provided an abridged response. Nonetheless, all of the participants’ voices and views are represented in this study. The primary themes that emerged from the data gathered from research question # 2, and the associated sub-questions are displayed in table 5.
Table 5.
*Themes in Response to Research Question 2.*

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Increased support / feedback from administration and colleagues</td>
</tr>
<tr>
<td>2</td>
<td>Trust in leadership</td>
</tr>
<tr>
<td>3</td>
<td>Accountability increased due to DDM’s</td>
</tr>
</tbody>
</table>

**Increased Support / Feedback From Administration and Colleagues**

Support was the most frequently identified theme evoked from question # 2. Each participant voluntarily stated that they appreciate a very positive relationship with their supervisor and valued his support throughout the new teacher evaluation process. However, few teachers indicated a change in their relationship due to the new evaluation program. Nonetheless, almost all participants reported an increased level of support from their supervisor. Most participants cited the formative reports sent from their supervisor after the mini-observations as valuable feedback; support that was not present with the previous evaluation process. In most instances, participants used the words support and feedback synonymously.

Susan stated that she felt that the new evaluation process was pushing her to perform at a higher level. She contended that her supervisor’s ongoing support increased her motivation as a physical educator. Susan stated: “I was commended for posting my lesson agendas and the mastery objectives for the start of each class. That encouraged me to keep doing the extras that this requires.” Barbie recognized a definite change in the relationship with her supervisor, but specified it was a positive and strictly professional change, as they have always gotten along very well personally. She was quick to state that her supervisor has always been consistently
supportive even though she was only observed and received feedback twice a year every other year. Barbie believes the new system is more meaningful because she receives the regular ongoing feedback from the written evaluations based on the mini observations. She appreciates that her supervisor visits more regularly and as a result offers more feedback, both commending and recommending. Middle school teacher, Jim, agrees with Barbie by contending that supervision is now better because their supervisor is seeing more. Jim thinks he receives “much more insightful feedback” as a result of the new system. Jim credits the more accurate feedback to the fact that these are mostly unannounced visits. Bill, another middle school teacher, also agrees that feedback is more frequent and credits a higher degree of communication as a result of the new evaluation system.

Larry indicated that his professional relationship has changed minimally and cites the feedback as “more complex.” When asked to expound on his statement, Larry commented that the information that he now receives has been more specific than he received from the past evaluation process. When asked if he found this useful, he stated further “it was definitely helpful and mostly positive.” David stated that when he came to this district it was at the beginning of all the changes in the evaluation system. Therefore he can’t provide a perception regarding a change with his current supervisor. However, David did state that even though in his previous position he shared the office with his supervisor, he receives “much more feedback” from his current supervisor despite only six mini-observations a year. Liz states: “I really like that my supervisor comes in often and observes my lesson then writes it up online and I can read the feedback and then plan accordingly.”

The participants also reported that they believed that their supervisor did an excellent job in procuring high quality professional development as an effort to roll out the DDM process
smoothly. Danielle stated, “I think our supervisor sought out the local experts in the field on assessment in physical education and provided us with some great ideas to get this thing (DDM’s) going.” Danielle’s colleagues echoed similar thoughts numerous times during the one-on-one interviews.

Veteran teachers indicated that they felt supported by their younger colleagues during the development and incorporation of DDM’s. Veteran elementary teacher, Larry, stated about one younger colleague: “…Mary has really helped everybody to understand what to do.” Twenty year secondary teacher, Rob, said about three younger colleagues: “…we’re seeing a new system now. It’s not the same as when I was in school. But for Mary, Leo and Liz it’s different. They know what to do.”

**Trust in Leadership**

Trust in their supervisor was a theme that was cited by a majority of the participants and clearly had a positive impact on the DDM implementation process. When questioned by the researcher about their relationship with their supervisor, there was an unequivocal tenor of trust asserted by all participants.

Danielle’s comments were representative of the majority of the staff. She explained, “I think our director did a really good job trying to prepare us… he is conscientious and I trust that he had good intentions with bad information.” Larry stated “I do whatever our supervisor tells me to do… I have always trusted him… he knows what he is doing. He’s always ahead of the curve.” Joe believes “his expectations are high for everyone, and fair.” Middle school physical educator, Will, states: “I’m glad he supervises all of us now…he knows what he is doing and I trust he will be fair.” Liz commented: “I believe that anything he tells us is accurate. He is always ahead of the curve.” Barbie “thinks our supervisor has been more than fair.” She also
eluded that she trusts his insight on the new educator evaluation program. One other secondary physical educator, Rob, sums up his level of trust with his supervisor as follows: “You know… he’s around all the time…he always sees what we’re doing…it doesn’t make him any more authoritative if he’s holding a red folder.” Liz believes that the level of accountability has definitely been heightened due to the Student Impact Rating. She contends that when her supervisor comes in for a mini observation, he is looking for specific teacher behaviors and whether or not her students are learning. She explains: “…when he gets back to me he is saying ‘This group of students were getting it, but not so much this group’ and he’s giving me suggestions on how to make it work.”

Almost all of the participants talked at length about their trust in their supervisor, who has led the department for fifteen years. Their supervisor also taught physical education, at every level in the district, for twenty years previous to becoming the Director of Athletics and Physical Education.

**Accountability Increased Due to DDM’s**

Another theme that was common among the participants was accountability. This theme typically pertained to the expectations of teachers as a result of the new educator evaluation program. The theme particularly resonated with the related data collection and the analysis as it pertained to the use of District Determined Measures in an effort to establish a Student Impact Rating for all teachers.

When asked about their supervisor’s expectation around the use of data, derived from the measure of student growth of students, the majority of teachers indicated that teacher accountability was a protuberant expectation. While some teachers expressed anxiety over the
level of accountability due to the new Student Impact Rating, the majority found it a reasonable expectation. Elementary teacher, Joe, expressed his sentiments on accountability as follows:

I think I’m fine with it (accountability)… I think it’s a quite reasonable expectation for administrators to expect to see (student) growth. I also think that it’s not a hard standard to achieve for the teacher…they’re not telling us what the DDM is. If they were telling us, “well here is your DDM. Make sure seventy-five percent of your third graders attain it.” That would be quite concerning to me. We get to make that decision. I’m fine with it and like it.

Barbie believes that this new administrative expectation can actually enhance the status of physical education within the content areas and even elevate it to a core subject similar to the way that fine arts are viewed. She states: “Yes, I think that’s the key (accountability) in their push to making our profession relevant.” Will is perceived by many of his colleagues the likely successor to their current supervisor and offers a similar perspective to Barbie’s on accountability. He states: “I like this (DDM’s) because it holds us accountable. Anytime we hold ourselves accountable it is a very good thing.” Susan elaborates further on the theme of accountability. She comments:

DDM’s…? I think they are used basically to hold us accountable… I like it and I don’t like it…it has definitely taken some fun out of teaching. But, I want to say it is a change for the better, this requirement for tracking student growth makes me feel held accountable to my students, and it’s about the students…right?

Danielle believes that DDM’s are all about accountability. She contends that it is her responsibility to have students learn. Danielle believes: “If DDM’s can show me, my students, my bosses, and the state that students learn in my class, then I am all on board with this new level of accountability.”

Lee makes sense about accountability as follows: “We’re here for the students. How effective we are plays into their educational experience…keeping teachers accountable in varied
ways is important…you can lose years in terms of their (students') education without accountability.”

Summary of Findings

In summary, the data gathered from question one is as follows:

- **Anxiety:** the majority of physical educators were anxious, to varied degrees, around the use of student growth to measure teacher effectiveness. The participants used terms such as “nervous”, “stress”, “pressure” and “scared” to express these varied levels of anxiety. The particular factors that produced anxiety were varied and tended to differ by grade level. Consistent factors that led to anxiety were the initial administrative rollout of DDM’s and the process used to decide the varied levels of student growth (low, moderate and high). Elementary physical educators expressed anxiety around the minimal physical education contact time with students. They felt that one class period per week for sixty minutes did not afford the amount of time needed to demonstrate student growth. This circumstance may not be unique to the research study site and is likely a universal issue at the elementary level across the state. However, there is potential for this concern to be present at the secondary level in districts that provide minimal physical education opportunities. Secondary level teachers were concerned that an influx of new students into the district, that are English language learners, could prove problematic with one of the DDM’s that was chosen that placed an emphasis on improving student communication, collaboration, creativity and critical thinking through adventure based education experiences. However, this concern may be unique to this specific research site. Ultimately, those physical educators that were recent graduates, less than ten years,
from a physical education preparation program or PETE expressed less anxiety than their veteran peers with ten years or more removed from their PETE program.

- **Preparation:** The theme of preparation was expressed in three contexts. The most prevalent context was directed towards higher education teacher training programs. Once again, there was a variance expressed by physical educators around this sub-theme. Similar to the theme of anxiety, those physical educators that were trained within the last ten years felt that they had been provided the expertise to implement the type of performance based assessments required for implementing DDM’s. Those physical educators that had been trained by physical education teacher education programs greater than ten years ago were not as confident about the assessment process required to demonstrate meaningful student growth. However, regardless of the training, it was atypical for any of the physical educators to have actually utilized performance-based assessments in their practices prior to the implementation of DDM’s. The second most mentioned sub-theme around preparation was directed towards the level of professional development provided by the research site. The majority of teachers felt that while the professional development was adequate, there was much confusion about what the state required around the process of implementing and utilizing District Determined Measures. Finally, several teachers expressed the significance of lesson preparation as a factor around the use of DDM’s. Some teachers felt that the use of DDM’s created a need to regularly incorporate opportunities for students to practice the selected DDM’s after the pre-assessment and leading into the post assessments. As a result, it was expressed that lesson preparation was impacted.
• **Need for change:** Initially, most participants did not report a need to change the evaluation process. Ultimately, most participants indicated a belief that the new evaluation program had created change that was positive to their pedagogical practice and had potential to elevate their profession in general. Ten participants cited specific positive changes in their professional practice once the new evaluation program had been implemented. These changes occurred in their lesson preparation and the inclusion of new student assessment strategies. These physical educators specifically cited the implementation of DDM’s as the major catalyst in the changes made to their pedagogical practices. The remaining six participants, that stated the new evaluation program has not prompted change, did in fact cite changes in their practices in follow-up questions throughout the interviews. The majority of the participants also expressed the sentiment that, while perhaps unintended, the student growth data derived from DDM’s could perhaps elevate the status of the content area of physical education. As a result, physical education could perhaps be recognized as a core curriculum such as art and music are categorized by NCLB. In summary the data gathered from question two is as follows:

• **Support:** The theme of support appeared in responses to both questions, but it was most prominent in question two. As previously stated, participants often used the term support and feedback interchangeably. Few teachers indicated a change in their relationship due to the new evaluation program. Those few teachers that did recognize a change in their relationship specified an increased communication around a new level of support through feedback as the variation. However, almost all participants reported an increased level of relevant feedback from their supervisor, therefore inferring that a change in their relationship may have occurred. A majority of the participants cited they appreciated the
support / feedback provided via the formative reports sent from their supervisor after the mini-observations. Some physical educators indicated that it was also common for their supervisor to provide words of encouragement at the conclusion of a mini-observation or when reviewing the data derived from their respective DDM’s. All participants that cited an increased level of feedback believed that it was more relevant and specific as opposed to the more generic level of feedback provided with the old evaluation process. Several of the physical education teachers credit a higher degree of communication as a result of the new evaluation system. It was clearly stated by the teachers that they felt supported by their supervisor throughout the transition to the new educator evaluation program. They were particularly appreciative of the increased level of feedback that their supervisor was now able to provide. In addition, veteran teachers believed that they felt support from their younger colleagues during the entire process of developing and designing DDM’s. They specifically appreciated the patience and confidence their younger colleagues displayed in this process.

- **Trust:** A common theme collected from question two, and the related sub-questions, was trust in leadership. The vast majority of teachers cited a superior level of trust that they had in their supervisor. Despite numerous statements that represented a strong level of disappointment in the roll out of the DDM process, there was an equally strong level of agreement that their supervisor was not the cause. In fact, there was strong consensus that their supervisor was doing his best to provide a smooth introduction and implementation of the DDM process. Participants spoke at length about the level of confidence that they have in their supervisor. However, many of the participants
expressed dissatisfaction with the process that Massachusetts Department of Secondary and Elementary Education introduced the concept of DDM’s.

- **Accountability**: Every participant, in responding to either question one or question two, mentioned the theme of accountability. However, the theme of accountability was most prevalent in question two. The vast majority of participants believed that their supervisor was holding them more accountable for their performance due to the new evaluation system. Participants cited the increased number of observations, as well as the required demonstration of student growth, as the key stimuli for increased accountability. In particular, teachers cited the collection and analysis of student growth data derived from the measurements of their specific DDM’s. An unexpected sentiment by the researcher was the strong agreement by the participants that the higher level of accountability was a positive of the new evaluation system. Several teachers expressed their desire to use the data demonstrated by the measurement of student growth as a potential benefit to their profession and individual practice.

**Chapter V: Discussion of the Findings**

This chapter will review the problem of practice, the methodology utilized to conduct this study, and examine of the major findings from the research. Additionally, the findings are presented within the context of the theoretical framework and the literature review. There will also be a discussion of the implications for practice, the limitations of the study, validity of the study, and conclusions. Finally, recommendations for future research will be presented.

**Problem of Practice**

As a condition for committing to participate in the “Race to the Top” initiative, the Massachusetts Department of Elementary and Secondary Education required participating school
districts to implement a new educator evaluation program. One element of the new evaluation program required districts to design and implement District Determined Measures that were intended to gauge a teacher’s impact on student learning. This approach to teacher evaluation has also been known as “Value Added” teacher evaluation.

The ongoing problem of this RTTT initiative, and the research problem of this study, is physical education teachers are being asked to incorporate a performance based assessment practice when they have little experience in this approach to assessment.

Because many physical educators used grading criteria that were often derived from a student’s attitude, effort, and proper attire for class incorporating an assessment process that is performance based may become problematic. Melorgrano (2007) contends that students in a physical education class rarely benefit from assessment practices that are subjective and based merely on compliance. Graham (2008) contends that a standards based approach to grading, in physical education, is the most valuable measurement of teacher effectiveness. Compliance based grading are difficult to quantify and cannot be identified with a standards-based approach. Nonetheless, many physical educators continue to use this approach (Melorgrano, 2007).

Given the likelihood that physical education teachers do not employ a performance based assessment process, there is potential for resistance from current physical educators. Accordingly, a research study to gain the perspectives of physical educators around the incorporation of DDM’s may prove beneficial to all stakeholders

**Review of Methodology**

The purpose of this study was to understand the perceptions of physical education teachers around the implementation of District Determined Measures as a tool to measure teacher effectiveness. This study was specifically designed to address the following questions:
1. How has the new educator evaluation process that now includes District-Determined Measures effected change in the physical educators’ instructional practices as perceived by the physical education teacher?

2. How do physical education teachers perceive any change in their professional relationship with their administrator due to the new evaluation process that now includes District Determined Measures?

To answer these questions, a qualitative bounded case study was designed and conducted with 16 physical education teachers employed in a large socially and economically diverse suburban school district outside of Boston, Massachusetts. The approved site consisted of nine schools with a growing enrollment of approximately 5,300 students (MA DESE, 2015). Data was collected through in depth, unstructured interviews, document review, and observations.

**Discussion of Major Findings**

The findings of this study were derived from the careful analysis of data primarily derived from semi-structured, in-depth one-on-one interviews. Additional sources of research data used to support the findings, included pertinent in-district memos, minutes of department meetings, and field notes accrued through observations of relevant department meetings. Additional sources of data enable findings to be triangulated in order to ensure the findings of a study are reliable (Creswell, 2007).

Through careful analysis of the research data, and two levels of coding, the following themes emerged:

1. **Anxiety** was expressed by a majority of the participants over the changes to the current educator evaluation program. In particular, the participants felt an increased level of accountability as well as pressure to demonstrate student growth. Additionally, some
participants expressed concerns that the process was rushed, with information presented at one meeting often being later revised at a subsequent meeting. This resulted in confusion and angst. However, the physical education staff made clear in the interviews that they held the Massachusetts Department of Elementary and Secondary Education and central office accountable for the confusion, not the direct supervisor.

2. **Preparation** for the new mandate was discussed in three contexts. The most salient context was directed towards the level of preparation provided to physical education teachers in their undergraduate and graduate level physical education teacher education programs necessary to implement the performance-based assessments required for DDM’s. Participants voiced mixed feelings around their undergraduate or graduate level course work. Participants that had graduated or completed a program within the last ten years felt capable and willing to implement the performance based assessments, rather than compliance grading. An interesting anomaly around these sentiments was despite their confidence and eagerness to apply performance based assessments, the recent graduates were not utilizing this assessment practice prior to recent expectation. In fact, they reported incorporating the compliance based grading systems and assessment practices that were in place when they began their teaching career. Veteran teachers were emphatic that they had never received the necessary training to implement performance based grading either in their undergraduate or graduate teacher preparation programs.

Another contextual use of the preparation theme was related to the school district based professional development provided to the physical education staff specific to District Determined Measures. Sentiments around this theme were also divided and varied by elementary and secondary teachers. Based on research findings, elementary
physical educators expressed satisfaction with the district wide professional development in contrast to secondary teachers who expressed disappointment with professional development.

The theme of preparation was also apparent within the context of lesson planning. Many of the participants cited an increased amount of lesson preparation for their daily instruction, although a variance in the preparation did exist by grade level. For instance, elementary teachers cited employing numerous opportunities for students to regularly practice the specific DDM after the pre-assessment and leading into the post assessment. Two elementary teachers correlated this action with “teaching to the test.” Secondary level teachers indicated greater research about the practice of incorporating formative assessments. They specified an increase in the regular use of formative assessments in order to check for learning. The use of formative assessments was specific to the DDM, but secondary teachers did indicate that it had expanded their repertoire of assessment strategies.

3. **The Need to change** was another prevalent theme. When prompted, the participants were quick to state that prior to the implementation of the new educator evaluation program, they did not recognize a need to change to the new evaluation method. Nonetheless, after the implementation of the new evaluation program, most participants recognized a positive change in their professional practices. Also, a majority of the participants agreed that the addition of DDM’s to the evaluation process provided impetus for this change in their pedagogical practice.

4. **Accountability** was also a common theme. Regardless of the number of years since the completion of undergraduate or graduate course work or grade level taught, numerous
teachers spoke to the increased level of accountability that resulted from the DDM element of new educator evaluation program. However, recent graduates were less apprehensive about the increased level of accountability. Despite a higher level of apprehension articulated by veteran teachers around incorporating performance based assessment practices, veterans participated fully in the development and implementation of DDM’s.

5. **Support** was a theme that participants specifically related to the increased level of contact and professional discourse that occurred with their colleagues and supervisor as a result of the new evaluation program. Most participants cited the increased amount of relevant feedback from their supervisor as a result of the mini observations. They felt this ongoing and regular feedback provided support during a stressful time in their careers. Additionally, most veteran teachers felt support from their younger colleagues during the development of DDM’s. The veteran staff appreciated their younger colleagues insight and confidence about incorporating DDM’s. Overall, most participants indicated that they now valued the new evaluation process more so than the previous system of teacher evaluations. They believed the increased support and feedback they received from their supervisor and the collegial support was a very valuable outcome of the new evaluation process.

6. **Trust** was a theme expressed by the majority of the participants. Participants specifically cited unwavering trust in their curriculum leader. Many of the participants spoke at length about their level of trust in their supervisor, who has led the department for fifteen years. Because of this trust, the participants were willing to comply with the changes necessary to implement the new evaluation process.
Findings in Relation to the Theoretical Framework

The Race to the Top initiative has produced far-reaching changes in public education. Changes such as establishing common core curricular with a new high stakes testing format, and a new approach to educator evaluation (Race to the Top, 2013) have been the most prominent. The most significant change for untested subjects, especially physical education, has been the aspect of the teacher evaluation program that measures teacher effectiveness based on student performance. (Holt-Hale et al, 2012). The success of these latest reform efforts remains to be seen. Sararson (1991) contends that chronic failure of school reform has had less to do with the ideas and more to do with the process of change. Therefore, the researcher viewed this study through the lens of educational change theory. Lewin’s Three Step Change and Prochaska and DiClemente’s Change Theory provided a framework in which to research the perceptions of physical educators around a change to a new teacher evaluation program that includes DDM’s. These two theories were selected based on the writings of Michael Fullan and John Kotter, two of the most highly regarded scholars and authors on the process of change. Kotter views the topic of change from the business perspective. Fullan studies change in the genre of education. Their collective wisdom, on the process of implementing change, points to the significance of collaboration between the agents of change and the targets of change. Lewin’s Theory views change from the perspective of the change agent, while Prochaska and DiClemente’s Change Theory looks at change from the target of change.

The Findings within Lewin’s Three-step Theory. Lewin’s there-stage model offers many useful insights into the process of change. This change effort can be conceptualized as a three-step process, where the targets of change are first prepared for the change, then change is
implemented and finally the new behavior becomes permanent (Burnes, 2004). This three-step process is often referred to as *unfreeze, change* and *re-freeze*.

Lewin contends that many change efforts fail because people are not prepared for the change (Burnes, 2004). When employees are not prepared for change, there is likely resistance and they are less likely to function effectively under a new system. Lewin believes that communicating the change well in advance to the employees will help them to feel more comfortable (Burnes, 2004). Research shows that those who have more complete information about upcoming changes are more committed to the change (Wanberg & Banas, 2000). Furthermore, in successful changes, the leader provides an overall vision for the change. (Herold, Fedor, Caldwell, & Liu, 2008). These ideas are supported by findings in this study. For example, the supervisor provided his staff with information about the implementation of DDM’s, long before other departments were notified. Many of the participants believed that they were ahead of the curve and as a result were ready and comfortable with implementing DDM’s. Furthermore, most participants believed that the professional development they received in preparation for this change was of high quality. However, some participants cited misinformation from the central administration, and ultimately DESE, and felt the change process was somewhat rushed. However, those perceptions were limited to secondary physical educators.

Another aspect of step one in Lewin’s Theory of Change, unfreezing the employees, is to build a coalition of believers that recognize this change could be beneficial (Burnes, 2004). In other words, expecting everyone to buy into the change at the same time may be an unreasonable task for the leader. Understanding the social network within the staff is crucial in order to identify leaders that can agree that change is needed, and as a result may be able to sway those
not convinced to get on board (Herold et al, 2008). Once again, there is evidence from the research data that a solid contingent of staff were supporters of the initiative. These supporters believed that this change (assessing students using DDM’s) has the potential to produce data that could lend more credibility to an undervalued curriculum area, evidenced by NCLB’s exclusion of physical education as a core content area. Additionally, the success of this process can be supported by the consideration there were no outright dissenters of this initiative as based upon the data collected through the interviews, observations, and relevant documents.

Providing support and allowing employees to participate in the process of change is also a vital aspect of unfreezing the current culture (Burnes, 2004). Lewin believes that management must listen to concerns of the employees and respond through the demonstration of emotional or instrumental support (Herold et al, 2008). Findings of this study also support these ideas, as numerous participants credited their supervisor with providing strong support throughout the piloting phase of the change process. In addition, the participants were given an opportunity to collaborate and select the DDM’s for their grade levels. Some participants pointed to this aspect of the change process as the key to their willingness to change. They felt that being allowed to voice their opinions and ultimately collaborate with colleagues to select a DDM, (as opposed to being handed a DDM) provided a more positive attitude towards the change.

Step two of Lewin’s theory is executing the change. Burnes (2004) states that the success of implementing this phase of Lewin’s theory may depend upon the type of change. He provides three tips that may assist to facilitate the change as follows: 1) continue to provide support, 2) create small wins and 3) eliminate obstacles.

There are findings in this study to support the theoretical alignment with this phase as follows: 1) participants regularly cited the support that they were provided by their supervisor
throughout the process. They particularly noted the regular and relevant feedback that was provided, 2) creating small wins can be aligned with allowing the teachers to select, design and pilot the chosen DDM, 3) In review of the grade level meeting notes, there were a few potential obstacles cited by the physical educators. They commonly expressed apprehension about maintaining a safe and productive classroom environment while the individual or small group skill assessments were conducted. The teachers cited the potential for off-task student behaviors while they conducted individual or small group assessments. They were also concerned about the accuracy of scoring student assessments during class time. The supervisor recognized this concern as reasonable, and as a result provided each physical educator an iPad mini to record students while they performed the specific skill or action. The use of video would allow teachers the opportunity to score students outside of class time. Additionally, the teachers would have more time to reflect on a students’ score thereby increasing validity and reliability in the process. In fact, the PE Metrics assessment tool recommends the use of an iPad. As a result, his teachers were able to maintain lesson momentum, minimize any unnecessary downtime with the use of this technology, and follow the PE Metric protocols when scoring.

Refreezing, the third and final step of Lewin’s model occurs once the change has been recognized as a part of the organizational culture (Herold et al, 2008). The change is considered successful if it has revised the ways of thinking and if the desired behavior change has become routine in the organization (Burnes, 2004). Once recognized as successful, management can reinforce the refreeze by publicizing the success, rewarding change adoption, and embracing continuous change (Gale, 2003).

While there is evidence that the change in the evaluation process, and in particular the adoption of DDM’s, has been accepted by most of the participants, it is difficult to determine if it
a permanent behavior change. Also, there is new information from the Department of Education that there may be revisions to DDM’s (personal conversation, Massachusetts Department of Elementary and Secondary Education, May 14, 2015). However, while it is only speculation that the changes to the Student Impact Rating will affect untested subjects, it is likewise too soon to classify the adoption of DDM’s as permanent. In fact, at the final secondary department meetings there was discussion among the physical educators that they may revise the current DDM’s. Accordingly, while progress has occurred, it is too soon to state that change is imbedded within the culture of the department.

**Findings related to Prochaska and DiClemente’s Change Theory.** Prochaska and DiClemente’s change theory, also recognized as the Transtheoretical Model or TTM, is considered a cyclical model of change. This model of change describes six stages of the change process that exists within the target of change. TTM, created by Prochaska and DiClemente in 1983 was typically applied to modify health behaviors. Nonetheless, TTM is considered one of the more popular change models and is used in a variety of change contexts. TTM is recognized as a dynamic model, because it takes into account behavioral setbacks (Prochaska & DiClemente, 1983). This spiral model suggests that the target learns from their relapses instead of circling around the issue, and as a result may enter the model in any one of the first five stages. The six stages include: 1) During the *pre-contemplation stage* phase the target has no intention of changing yet and is unaware of or denies the need to change, 2) In the *contemplation stage* the target has recognized that there may be a problem and change is needed but the process has not yet begun, 3) The *preparation stage* is when the target begins to focus on what can be done and on developing a plan for change, 4) In the *action stage* the target publicly states a desire to change, enlists help to make the change, and sets realistic goals for the change, 5) The
The *maintenance stage* occurs after the change is in place and the target is attempting to prevent any relapse into old habits or behaviors, 6) *Termination* occurs when the change is ingrained and ongoing (Burnes, 2008).

Prochaska and DiClemente’s change theory was chosen by the researcher as a lens in which to view this study specifically because it places the initiative to change behavior on the target of change. However, the findings in this study do not smoothly align with this theory. The findings that had mixed results are as follows:

1. **Pre-contemplation:** This stage may have existed initially. The findings show that the participants were quick to state that prior to the implementation of the new educator evaluation program, they did not recognize a need to change to the new evaluation method. Ultimately, the vast majority of the participants cited an appreciation of the new evaluation process and the use of DDM’s.

2. **Contemplation:** Once introduced to the new evaluation process, the participants recognized that change was inevitable. In most instances, the participants expressed, in their interviews, that they recognized the previous method to evaluate teachers provided minimal relevant feedback.

3. **Preparation:** The supervisor introduced the Student Impact Rating aspect of the evaluation process and then provided professional training that was aimed at the development of DDM’s.

4. **Action:** Based upon participant interviews, it is difficult to determine an alignment with this stage based upon the research findings. The findings point towards the supervisor providing the motivation for this change, but not necessarily the targets themselves. The supervisor also enlists support from experts from higher education in the field of
assessment in physical education. These experts introduced tools, such as the PE Metrics Assessment package, along with suggested timelines for the implementation of DDM’s.

5. Maintenance: The participants of this study will likely enter this stage in the upcoming school year after their revisions.

6. Termination: Since this stage occurs once the change is ingrained, there were no findings in the research to support that this final stage of change was attained during the four months of the study. Future research is needed to determine the outcome of this stage of the change process in the study.

In summary of the findings in relation to the theoretical framework, Lewin’s Three Step Change Model provided a very sound lens in which to view the study. In fact, this framework appeared to provide a model from which the supervisor could successfully implement change, with potential for replication in other change scenarios.

**Findings in Relation to the Literature**

The literature reviewed for this study surveyed teacher supervision and evaluation for the past three hundred years. There was a close examination of the process within the past 30 years, and in particular the effect of Value Added Modeling as an element in the process of teacher evaluation. This narrative will reference the most salient points of the literature review and align those points with the findings within the study.

Value Added Modeling is one source of teacher performance data that has the potential to provide valuable insight around teacher effectiveness. The use of Value Added Measures is intended to isolate the contribution each teacher makes in that particular school year toward his or her students’ growth based upon a pre-determined measurement or standard. The incorporation of Value Added Modeling, also known as VAM, as a source of performance data
used to determine teacher effectiveness is controversial (Baker et al, 2010). Nonetheless, many noted scholars of education such as Charlotte Danielson (2010), Linda Darling-Hammond (2011) and Robert Marzano (2011) believe that if used with other sources of data, VAM can be a valuable resource when determining teacher effectiveness. In the new teacher evaluation program for Massachusetts, VAM is termed as the “Student Impact Rating” and is decided by the use of District Determined Measures. As a result, the DDM’s serve as an additional source of evaluation data for all Massachusetts teachers and administrators.

The Brown Center on Education Policy of the Brookings Institute sponsored a study on the use of VAM in teacher evaluations. This study strongly advocated for the use of VAM in teacher evaluation. The study contended that the quantitative data derived through VAM could be useful to teachers (Glazerman, 2010). This contention is supported by the findings of this research study as participants have reported that the data collected through the use of DDM’s is useful. Barbie, a high School physical educator, provides a succinct sentiment referenced by several participants: “What I think is most valuable is looking at the data….and saying ‘how do we use this (data) to make us better’ and then collaborating with colleagues to do it.”

Critics of VAM point out that the use of VAM to evaluate teachers has not yet been scientifically validated. Newton et al, (2010) believe that numerous factors can influence a students’ academic growth. They cite in-school factors as well as out of school factors. Participants of this study from all grade levels expressed similar concerns. David, a novice physical educator, expressed his concern around his evaluation being predicated on student growth. He believes that the minimal weekly contact time with his students is a problem difficult to overcome. He also alluded to the fact that recess time is minimal and a student may not have the same exposure to physical activity experiences outside of school. He believes that
these factors can also contribute to a lack of student growth. Lisa, another elementary physical educator provided a stronger statement regarding in-school factors negatively effecting student growth as follows:

You have them once a week. If they were in a math class, are you going to teach them the lesson on Friday and then have them come in the next Friday and expect them to remember what they’ve learned and be able to build upon that? To me that’s a problem.

Danielle, a veteran high school teacher, further elaborated these concerns. She stated that she has numerous students that are English language learners and new to the country. As a result, Danielle is concerned that the language barrier and lack of prior experience may reflect poorly on her ability to demonstrate student growth.

In his Carnegie Foundation sponsored study, Douglas Harris (2012) states that it is critical that these value added measures are based upon valid and reliable criteria. Validity refers to the degree to which something measures what it claims to measure. Reliability refers to the degree to which the measure is consistent when repeated. This consideration also aligns with the findings of this research study, but in a contradictory manner. Based upon a personal conversation with a member of the Educator Evaluation team from the Massachusetts Department of Elementary and Secondary Education (November 17, 2014), the use of a valid and reliable assessment tool is not required. Nonetheless, the supervisor did provide his staff with PE Metrics, a well known and highly regarded valid and reliable assessment package for physical education.

The role of the leadership in the implementation of a new evaluation process is significant and can make or break the process and cannot be understated (Halverson, Kelley & Kimball, 2004). The participants in this study demonstrated a high level of confidence in the leadership provided by their supervisor. Veteran teacher, Rob, provided a comment similar to
his colleagues’ beliefs. He speaks about his supervisor regarding the implementation of DDM’s:
“ I think he has done a fantastic job. I think he keeps up with the standards, probably on the
forefront of most high schools in the state.” Danielle, another veteran, agreed and stated: “I
think our supervisor sought out the local experts in the field on assessment in physical education
and provided us with some great ideas to get this thing (DDM’s) going.”

Lewin (as cited in Covey, 1990) points out the importance of involving all stakeholders
in the procedure of creating change. When administrators involve teachers in the
implementation of a new teacher evaluation system, it is more likely that they will be committed
to inventing solutions and also likely to accelerate effective implementation (Sergiovanni, 2001).
Findings within this study support Sergiovanni’s (2001) contention. Many of the participants in
this study valued the opportunity to collaborate with colleagues to develop their DDM’s. Joe, a
veteran elementary physical educator stated it best: “they’re not telling us what the DDM is… If
they were…that would be quite concerning to me.” Barbie felt that by involving the staff in the
process there was a greater level of “buy-in” by all. She stated further: “I expected a lot of push
back when it (DDM roll out) was introduced. That was just the opposite.” These two statements,
although more lucid than other colleagues, were representative of most participants. Some
participants mentioned the complaints levied by their colleagues in core content areas that were
required to incorporate the results of MCAS as one DDM. Lee stated: “Even though our director
pointed out the benefit of PE Metrics, he didn’t jam it down our throats… we still could choose
something else.”

As described in the literature review, Danielson and McGreal (2000) explore the issue of
trust between the teacher and supervisor. They contend that many educators are more
knowledgeable about current best practices in instruction. They maintain that many
administrators, especially those whose background is not in the content area they are evaluating, would be hard pressed to spot content inaccuracies. This may be even more apparent within physical education, considering the perceptions that participants expressed in the interviews. As it pertains to findings in this study, there does exist a correlation as described by Marcia. In her interview, she was initially concerned that her director was not going to supervise her and one other colleague. Her new evaluator was an assistant principal, formerly a math teacher, and had little recent practical knowledge of physical education. She eluded that the revised curriculum and new instructional strategies they employ might be unfamiliar to her new evaluator. However, Marcia did feel relieved that her director and her new evaluator had communicated regarding what to expect in an observation, and as a result Marcia expressed a greater level of trust in the process.

Furthermore, Danielson and McGreal (2000) believe that a low level of trust existing between administrators and teachers can lead to a culture of passivity and protection among the teachers. The result may be a limited effort to change. If the opposite is true, that a high level of trust fosters change, then this contention is also supported by the findings, as all participants in this study expressed strong trust in their supervisor. They valued the fact their supervisor was a physical education teacher for many years prior to taking on his administrative role as Director of Physical Education and Athletics. They also mentioned his strong content knowledge and felt their department was always ahead of the curve as it pertained to current best practices in the field. As a result, participants reported a willingness to accept a controversial method of measuring teacher effectiveness.

In summary, the literature review concluded that the use of Value Added Measures as a tool to evaluate teacher effectiveness was a useful and worthwhile indicator of teacher
effectiveness if combined with other sources of data. In the study, *Problems with the Use of Student Test Scores to Evaluate Teachers*, Baker et al (2010) summarized their research and concluded that student test results should be included in the teacher evaluation process, but should only be one piece of information used to evaluate effectiveness. The application of VAM with the new educator evaluation program in Massachusetts followed Baker’s et al (2010) contention.

**Limitations**

There were limitations to the study. The ability to generalize the findings may be limited due to several factors. The results were based upon the perceptions of only sixteen physical educators within one school district. The school district is well funded and therefore able to provide high quality professional development. Additionally, my known familiarity with the direct supervisor of the participants may have hindered some responses. However, this should not automatically discount the reliability of the responses. The depth of the interviews and the consideration that the interview process was semi-structured allowed for follow-up questioning by the researcher as needed. As a result, the potential for this particular limitation may have been mitigated.

Despite the limitations, the findings of this study were positive as the majority of the participants perceived the use of District Determined Measures as useful for providing data that will demonstrate teacher effectiveness, improve instruction, as well as provide credibility for the content area of physical education. In particular, the participants cited changes in their practices around lesson preparation and the increased use of performance based assessments. Of particular note was the minimal amount of initiative pushback reported by the participants.
Conclusion

This study has demonstrated a willingness by the participants to accept this aspect of teacher evaluation, and it did so in alignment with the theories described in the literature review. Some of the literature review considerations that are clearly supported with the findings in this study include: trust in the supervisor / evaluator, involving stakeholders in the process of change, strong leadership in regards to implementation, and providing opportunities for collaboration amongst teachers around the quantitative data derived from the DDM’s to improve their practice.

In addition, the researcher believes that there was conclusive support for the use of Lewin’s Theory of Change as a theoretical framework in which to view the perceptions of the participants around the implementation of a new educator evaluation program that included the use of District Determined Measures. While the first two stages of Lewin’s Theory, unfreezing and change, were clearly presented by the participants, the final stage of re-freezing had not yet been vetted due to the timing of the study. More insight on this will be discussed in recommendations for future study.

The use of Prochaska and DiClemente’s Change Theory was not as conclusive. While there was evidence that the participants were aligned with some of the six stages of change, it did not appear to flow as cohesively as did Lewin’s Theory of Change. Prochaska and DiClemente’s theory, also recognized as the Transtheoretical Model or TTM, is cyclical and does allow for participants to enter any stage dependent upon their particular stage of change. Nonetheless, its’ application was not as fluid as was Lewin’s.

In chapter one, the researcher explored four possible theoretical frameworks from which to view the study. The two that were not selected included Lippet’s Seven Stages of Change and the Social Cognitive Theory of Change. In retrospect, the Social Cognitive Theory would have
fit well as a lens as it describes the behavior of the participants in two contexts. First, those recent graduates, who cited strong preparation from their university coursework, described a solid understanding and value on performance-based assessments, as used in DDM’s. However, upon entering the field, they succumbed to the compliance based assessment practices utilized by other veteran teachers for assessing students, a phenomenon described in the Social Cognitive Theory of Change. The second context was more subtle and eluded to one of the reasons the veterans were willing to accept the change in their evaluations. Throughout the interview process, the veteran participants often respectfully referred to their younger colleagues as ready for this change and expressed a willingness to follow them with the transformation. These behaviors reflect the environmental influences that can affect change, as discussed in the Social Cognitive Theory of Change.

Additionally, the researcher anticipated that teacher pushback might occur during the change to a new educator evaluation process that included DDMs. This premise was based upon the likelihood that veteran teachers would be ill prepared for such a drastic change in the assessment of students. However, based upon the findings, there was minimal evidence of veteran teacher pushback. The researcher can only theorize that this phenomenon was feasible due to the leadership that supported change. The supervisor appeared to follow Lewin’s Theory of Change as described in the findings based on the theoretical framework. As a result, the findings of the study appear to demonstrate a perception held by a large majority of the participants that the incorporation of DDM’s was viewed with an open mind, if not favorably.

Also, the participants in this study reported a positive change in teaching behaviors. Changes in teaching behaviors that were either reported by the participants or noted by the researcher through the interview process are included in the table below.
Table 6

*Changes in teaching practices as reported by participants or noted by the researcher.*

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Utilization of performance based assessment practices</td>
</tr>
<tr>
<td>2.</td>
<td>Development of goal oriented lesson plans</td>
</tr>
<tr>
<td>3.</td>
<td>Utilization of rubrics based on national standards to track student performance</td>
</tr>
<tr>
<td>4.</td>
<td>Increased use of technology i.e. heart rate monitors, accelerometers, related activity software / apps to track student performance</td>
</tr>
<tr>
<td>5.</td>
<td>Utilization of iPad to record video for the assessment of student performances</td>
</tr>
<tr>
<td>6.</td>
<td>Increased collaboration with colleagues</td>
</tr>
<tr>
<td>7.</td>
<td>Utilization of data collected through DDM’s to improve teaching practices</td>
</tr>
</tbody>
</table>

As cited by the physical education teachers that were recruited for this study, there were some conclusive findings, despite a distinct variance in their undergraduate or graduate level training. One significant finding was that a strong majority of participants indicated their support for the use of District Determined Measures as a source of data to determine teacher effectiveness. Through this study the researcher learned that physical educators that completed their professional training within the last ten years felt prepared and comfortable incorporating the performance based assessments that are a requisite for the incorporation of DDM’s. In contrast, those physical educators that had been trained in undergraduate or graduate programs beyond ten years felt less prepared and expressed a high level of anxiety around incorporating DDM’s. Nonetheless, the majority of the veteran teachers were willing to employ a performance based assessment practice in order to comply with the DDM requirements. In addition, the findings demonstrate that the participants reported favorable changes in their pedagogical
practices. These include an increased level of lesson and unit preparation, as well as an increase in the utilization of performance-based assessments.

With regard to experiencing a change in the relationship with their supervisor, the participants did not immediately recognize any change. However, there was evidence from a majority of the physical educators that they felt strongly supported throughout the changes in the evaluation process and appreciated a stronger level of feedback regarding their performance from their supervisor. Ultimately, the researcher contends that the findings show a positive change in the professional relationship between supervisor and the participants. However, there is no evidence to support that this positive relational change can be tied in to the incorporation of DDM’s, but may have been the result of the new educator evaluation process.

**Practical Implications**

Hanushek and Kain (2005) claim that teachers are the most influential factor in student achievement. Accordingly, it is critical that an effective teacher evaluation process is implemented in all schools. The evaluation program must utilize multiple sources of data to provide feedback in order to improve performance, as well as assist in the decision-making around teacher retention (Danielson, 2010).

This study was conducted to gain a broader understanding of how the use of District Determined Measures affects the practices of physical education teachers. While this research was conducted in a large, socio-economically diverse, suburban school district, there is potential for similar results to be found in other comparable districts within the state of Massachusetts or even nationally.

This study should be useful to those Massachusetts school districts, administrators and physical educators that will be piloting or implementing DDM’s as a data source to evaluate
teacher effectiveness as based upon student growth. First, it will be important for administrators to recognize that the comfort level of incorporating DDM’s will vary with staff. The administrators must also recognize that the variance will be dependent upon how recently the staff had graduated from an undergraduate or graduate Physical Education Teacher Education program. Secondly, it will be incumbent upon the administrator to provide high quality professional development in order to bridge the comfort levels of physical educators through the transition from compliance based assessment practices to the incorporation of performance based assessment practices, as required to meet the expectation of demonstrating student growth over a period of time.

Thorough planning is necessary to ensure that potential roadblocks are averted and that recognized assessment programs such as PE Metrics as well as the Tennessee Model are used as guidelines to enhance the process of developing and implementing District Determined Measures.

**Recommendations for Future Research**

There is an abundant amount of research available on the use of Value Added Modeling / Student Impact Ratings as an indication of teacher effectiveness. However, that research is extremely limited when directed towards grade levels or subjects that were previously untested. In a general context, the impact of DDM’s as it relates to content specialists such as art, music, physical education, health education, and primary grades teachers could be the subject of future studies.

Another study that may prove valuable could be directed towards the perceptions of supervisors / evaluators of physical education teachers. The administrators’ perceptions may or may not support the findings of this study particularly pertaining to favorable changes in
instructional practices as reported by the physical education teachers in this study. Nonetheless, such a study could prove beneficial with respect to the change process in education.

Finally, a study that would potentially extend the findings of this research could be directed towards the examination of the long-term changes in teaching practices of physical education teachers due to the implementation of District Determined Measures.
References


Doherty, J. (2007). *Perceptions of teachers and administrators regarding the implementation of a standards-based teacher evaluation process.* Retrieved April 7, 2014 from ProQuest


Massachusetts Department of Elementary and Secondary Education, Massachusetts Model for System for Educator Evaluation (2012). Retrieved from  
http://www.doe.mass.edu/edeval/model/


Race to the top. (2013). Retrieved November 01, 2013, from
http://www2.ed.gov/notclamped/programs/racetothebottom/index.html


August 27, 2010

educational change in Finland?. New York: Teachers College Press.


Seidman, I. (2006). *Interviewing as qualitative research: A guide for researchers in education*

Falmer.

Champaign, IL: Human Kinetics

Norwell, MA: Kluwer academic publishers.

Stake, R. (2005). *Qualitative case studies.* The Sage handbook of qualitative research (pp. 443


Thompson, E., (1952). So begins-so ends the supervisor’s day. *Educational leadership, 10*(2), 80-84.


