2015

Teacher Capacity and Attitude toward Data: An Examination of the Association between Teacher Beliefs and Student Performance on the Measures of Academic Progress Assessment

Elizabeth C. Mitcham
Gardner-Webb University

Teacher Capacity and Attitude toward Data: An Examination of the Association between Teacher Beliefs and Student Performance on the Measures of Academic Progress Assessment

By
Elizabeth C. Mitcham

A Dissertation Submitted to the
Gardner-Webb University School of Education
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Education

Gardner-Webb University
2015

Approval Page

This dissertation was submitted by Elizabeth C. Mitcham under the direction of the persons listed below. It was submitted to the Gardner-Webb University School of Education and approved in partial fulfillment of the requirements for the degree of Doctor of Education at Gardner-Webb University.

_____	_____
Kelly T. Clark, Ed.D. Committee Chair	Date
_____	_____
Jason Parker, Ed.D. Committee Member	Date
_____	_____
Jeffrey Hamilton, Ed.D. Committee Member	Date
_____	_____
Jeffrey Rogers, Ph.D. Dean of the Gayle Bolt Price School of Graduate Studies	Date

Acknowledgements

This dissertation has been the work of so many who have assisted me in other aspects of my life so that I could focus my attention on this work over the last year. I want to thank my parents – Dr. Glenn D. Crowe and Nancy W. Crowe, M. Ed. You raised me to believe in education and hard work. You have supported me and encouraged me and pushed me to be the best that I can be always. Your example of dedication to your professions has been an inspiration to me. Mom, thank you for babysitting, grocery shopping, cooking holiday meals at my house, and reminding me to find a balance even in this difficult pursuit. Dad, I could never have completed this process without your understanding of the struggle and constant words of affirmation and encouragement. Thank you for all the times you let me vent about my stress and my worries and always helping me see a way to move forward.

Jason and Drew, thank you seems way too minimal to say to you. You have sacrificed so much for me to pursue this dream. I appreciate the time and understanding more than I could ever express. Drew, I can only hope that you will be proud of me when you are old enough to understand. You have been more patient and more giving of your mom than any 5 year-old should have to be. I hope that seeing my hard work and dedication to a dream will inspire you to always follow your own with vigor and endless persistence.

To my extended family and friends, I appreciate your understanding of missed events, forgotten calls, and continued support over the last year. To my professors and colleagues at Gardner Webb University, thank you for sharing yourselves with me during the last 2 years. I know I will take messages from each of you as I continue in our field.

Dr. Sabrina Fordham, you have been my partner through this process at every step. I am grateful for the friendship that we share and the experiences we have built together. Your calm reassurance allowed me to survive. Dawn Farrington, this would not be without you and what you have brought to my life. My whole educational career has elements of your influence entwined in it. Thank you, dear friend, for being a steadfast encouragement!

Thank you to my committee members, Dr. Jason Parker and Dr. Jeff Hamilton, for your advice and the many great conversations that spurred me on when I was at an impasse. I am grateful for the opportunities I had to work with you. And to my chair, Dr. Kelly Clark, I cannot imagine a more perfect person to guide me through this process. I cherish the time we have spent together on this project, even when you were pushing me uphill. You knew exactly which words to choose each time to spur me onward. Thank you for your wisdom, your kindness, and your generosity with your time. I have learned so much from you. This work is the product of our vast number of conversations, and I am honored to have worked with you on this project. Your devotion to our craft is inspiring!

Lastly, I want to thank the community of educators who taught me, raised me, coached me, and helped guide my path through my career. I bring what I have learned from each of you to the education world every day.

Abstract

Teacher Capacity and Attitude toward Data: An Examination of the Association between Teacher Beliefs and Student Performance on the Measures of Academic Progress Assessment. Mitcham, Elizabeth C., 2015: Dissertation, Gardner-Webb University, Teacher Efficacy/Measures of Academic Progress/Student Performance/Teacher Beliefs

Since the onset of NCLB legislation and federal funding for schools tied to summative assessment performance, educational leaders have sought to identify factors that are most influential on student learning outcomes. Research continues to link the use of formative data practices and teacher efficacy to improved student performance. The intent of this study was to explore associations between teacher capacity and attitude and student performance. In this embedded mixed-methods study, qualitative data from focus groups comprised of survey participants were collected within a larger quantitative study that examined associations between teacher beliefs and student performance. The study focused on the impact of teacher beliefs on learning outcomes.

Participants included elementary math and/or reading teachers at four elementary schools in a large, urban school district. Perceptual survey data were collected regarding teacher beliefs about their capacity to use MAP data and their attitudes towards MAP data. These data were compared to student proficiency and growth scores obtained on MAP in both math and reading using Pearson product-moment correlation. Focus group data were collected from each site in order to explain trends in survey responses.

This study introduced teacher attitude as a new construct within teacher efficacy that was compared to both student proficiency and growth on the MAP assessment. Correlations for the relationships between teacher perceptions and student performance ranged from .24 to .46 (capacity) and .23 and .65 (attitude), with the highest correlational relationship between teacher attitude and student growth. Recommendations from the researcher include addition of teacher attitude as a separate construct within teacher efficacy and additional professional learning within the site of the study.

Table of Contents

Chapter 1: Introduction

As the results of high-stakes summative assessments have become a focus of educational and public policy industries, the need for resources devoted to predicting these outcomes has increased (Jones et al., 1999). Research suggests a strong link between formative data and summative student performance (Weiner & Hall, 2004), causing school districts to allocate increasing numbers of resources to formative data practices. Data-driven decision making (DDDM), applied to student achievement testing data, is a central focus of many school and district reform efforts in part because of federal and state test-based accountability policies (Marsh & Robyn 2006).

Since the start of educational reform initiatives such as No Child Left Behind (NCLB) and Race to the Top (RttT) that are directly linked to student outcomes on summative state assessments, there is an increasing need for states to adopt assessment practices that are rigorous and generalizable so that student growth can be measured (Wang, Jiao, & Zhang, 2013). Fiscal resources from RttT are based on educational performance outcomes related to student growth on standardized summative assessments (United States Department of Education [USDOE], 2008). Research conducted by Black and Wiliam (1998) concluded that classroom practices were formative and that when teachers use data collected from ongoing assessments of student performance, the decisions they make about instruction are likely to be better founded than those in the absence of that data. Connections between data-driven instructional practices and summative student performance launched the education trend of formative assessment as a standard instructional tool used by educators. As the need for this data in real time increased, computerized adaptive testing (CAT) that provided educators with immediate student performance data that were highly accurate and reliable increased in popularity

(Kingsbury & Weiss, 1983). These CATs provided multiple opportunities to assess students throughout a school year in a cost-effective, user-friendly platform (Way et al., 2010).

Problem Studied

Fiscal and human resources dedicated to student achievement are increasing each year in response to current educational legislation and funding resource criteria (National Center on Accessible Instructional Materials, n.d). According to the most recent Condition of Education report from the National Center for Educational Statistics, the U.S. per pupil expenditure is at the highest level in history (Kena et al., 2014). Even so, U.S. schools continue to be outperformed by other countries in proficiencies in science and mathematics, high school graduation per capita, and percentages of the public with advanced degrees (Gonzales et al., 2008). For this reason, identifying the factors that influence student performance on assessments is paramount. Heritage (2007) found that the skills necessary for solid formative assessment practices required teachers to demonstrate mastery-level teaching in their content areas and use of assessment. In a continuing study with colleagues, Heritage (2010) found that teachers were able to make generalized inferences about student achievement data but lacked the skills necessary to use the information to drive instructional planning. Schools and districts should engage their teachers in formative assessment practices and devote resources to ensure that the process is embedded in learning experiences with fidelity. It is necessary to provide structure and professional learning for teachers in using formative assessments to impact learning outcomes for students (Pinchok & Brandt, 2009). According to Howell, "The single greatest influence on learning is not socioeconomic status–it's instruction" (Personnel communication, June 2013). This statement highlights the rationale to focus

research on which teacher traits show the greatest associations with instructional practices that influence student performance. The current problem is that since teacher efficacy (TE) is an expansive construct (Guskey, 1982; Labone, 2004; Tschannen-Moran, Woolfolk Hoy, & Hoy, 1998), it is necessary to identify the components that impact student learning.

Theoretical Framework

The premise of TE is a basic concept with powerful impact. Central to this idea is a belief by teachers of their own abilities to provide educational experiences that result in the desired outcomes of student engagement and learning, regardless of external factors (Armor et al., 1976; Bandura, 1977). The theoretical framework of this study employed a postpositivist viewpoint steeped in the idea that human beings construct the reality around them based on internal factors and belief systems (Creswell, 2013). Efficacy promotes the idea that a person's beliefs about their ability to bring about an outcome (i.e., student achievement) are related to the outcome itself (Bandura, 2001; Parajas, 1996). Research has also indicated that personal beliefs are factors in outcomes related to student learning (Ashton & Webb, 1986; Bremer, 2008).

As a connection between these factors that impact student outcome and the current culture of assessment in schools, researchers have developed visual models that integrate the concept of efficacy, assessment practices, and student outcome measures (Balls, Eury, & King, 2011). In Balls et al. (2011), these models are referred to as value-added models that have been widely used to address changes needed in areas of student and teacher assessment. Balls et al. exposed the weakness to many models as lacking a system of feedback that is immediate.

Included in this text is a Value-Added Assessment Model created by Dr. Douglas

Eury, who offered this particular representation as a plan for "cultural transformation aimed at enhanced organizational performance" (Balls et al., 2011, p. 25). This system is based on the principles of efficacy in assessment practices that reference both dispositions and assessment skills similar to the constructs of teacher attitude and teacher capacity brought forth in this study.

Balls et al.'s (2011) large-scale Value-Added Assessment Model speaks to organizational systems that can be impacted by the efficacy of its members. Inherent in this model is an assessment practice that uses sustained routines and common language during professional inquiry with colleagues to make decisions that will lead to "measurable results in student learning" (Balls et al., 2011, p. 35). Masters (2013) offered a visual model of evaluation and feedback that supports the practice suggested by Balls et al. Masters' Educational Decision-Making Loop diagrams the process of educational decisions as acting on previous knowledge in a way that seeks to improve outcomes and ultimately life consequences. Though not stated in the same terms, Masters identified three components necessary to move a person or team to action that are congruent with the teacher capacity construct in this study. This study found strong, positive associations between teacher attitude (a construct not addressed in this model) and student performance outcomes, which indicate there are additional factors that impact action towards improved outcomes and life consequences while supporting the value-added assessment framework (Balls et al., 2011).

Deficiencies in Literature

Though much research has demonstrated a connection between TE and various instructional practices and researchers have continued to examine multiple constructs within the context of self-efficacy and teacher perceptions (Armor et al., 1976; Ashton &

Webb, 1986; Bandura, 1977; Tschannen-Moran et al., 1998), little is known about specific attributes of efficacy that are associated with student performance. Many studies have shown that teachers with high levels of efficacy regarding their beliefs about themselves and their colleagues impact student behaviors—both academic and behavioral (Gusky, 1982; Parjares, 1996; Rose & Medway, 1981; Woolfolk Hoy, Davis, & Pape, 2006). However, there is a call in the research for additional studies exploring the relationships between teacher outcome expectancies, personal efficacy (PE), and TE because there continues to be debate about which components of the complex construct are directly related to educational outcomes. Soodak and Podell (1996) called for continuing exploration of the dimensions of the efficacy construct related to the relationship between outcome expectancy (OE) and teacher behavior to validate some of the findings in their study. The distinction found in this study, supported by Bandura's (1997) notion of the differences in efficacy expectation and outcome efficacy, suggest that efforts to increase levels of efficacious behaviors should be focused on whether low TE is due to teacher lack of confidence in their skills (capacity) or a sense of futility regarding the impact of their work (attitude) (Bandura, 1997; Soodak & Podell, 1996). Bremer (2008) explored relationships between TE and student growth on Measures of Academic Progress (MAP) assessments and concluded that her findings corroborated that a relationship was present between these constructs. Bremer called for additional researchers to validate similar relationships between TE and MAP student performance data.

The current study attempted to answer the calls for additional research of Soodak and Podell (1996) by exploring relationships between specific attributes of efficacy and student performance as they relate to outcome expectations and PE and of Bremer's

(2008) recommendations for additional efficacy comparisons to student performance on the MAP assessment. Like Bremer, this study sought to inform the body of knowledge surrounding the factors about efficacy that contribute to increases in student performance outcomes on MAP assessments since the study district utilizes that data as a method of improving student performance on the state assessments.

Significance of the Study

Improving student learning outcomes is at the heart of all instructional practices (NCLB, 2002). Teachers spend countless hours in professional development in order to increase their capacity to provide quality instruction to students (USDOE, 2008). Accountability for student achievement has increased from the capital to the classroom (Weiner & Hall, 2004). Research continues to support that TE is an expectancy construct relevant to the understanding of how teacher sense of efficacy impacts student achievement (Ashton & Webb, 1986). Ashton and Webb (1986) went on to say that "[A] teacher's specific outcome expectations regarding the efficacy of teaching are filtered through their judgments of how able they are to influence student achievement" (p. 139). The significance of the problem studied lies in determining the associations of specific attributes of efficacy—a construct known to correlate with student achievement. Understanding the impact of capacity and attitude as related to efficacious teaching behaviors may inform educational decision makers on how best to focus their limited resources (Bandura, 1997; Bremer, 2008; Soodak & Podell, 1996).

Purpose of the Study

The intent of this study was to explore associations between teacher capacity and attitude and student performance. In this embedded mixed-methods study, qualitative data from focus groups comprised of survey participants were collected within a larger

quantitative study that examined associations between teacher beliefs and student performance. Collecting both types of data expanded the depth of the quantitative data and explored phenomena in the survey results by collecting qualitative perceptions from a cross section of participants (Creswell, 2013). Both types of quantitative data collected were aggregated by grade level; therefore, this mixed-methods design allowed for potential trends among individuals to emerge from the responses of the focus group participants. Hearing from specific participants also illuminated trends about teacher capacity and attitude that were not specifically measured by the survey and informed discussions and implications for future research.

Research Questions

The study sought to answer the following research questions in order to build on the theory of efficacy as it relates to teacher perceptions and student outcomes.

1. What is the association between teacher capacity in using MAP formative data and student performance?
2. What is the association between teacher attitude toward using MAP formative data and student performance?

Constructs in this Study

For the purposes of this study, the following constructs have been defined by the researcher based on trends and themes about specific attributes of TE and teacher beliefs found in the literature (Ashton & Webb, 1986; Bremer, 2008; Soodak & Podell, 1996).

Teacher capacity. The ability for teachers to access, understand, and utilize student data and instructional planning resources generated from MAP assessments.

Teacher attitude. Beliefs or feelings held by teachers about the use of MAP data as a valid formative measure of student performance and instructional decision making.

Student performance. This term is used in this study to describe overall student achievement using constructs of proficiency and growth as measured by MAP.

Basic Assumptions

This study assumes that the respondents of the survey self-reported honestly to the questions. It assumes that they have an accurate understanding of their abilities and feelings that they translated correctly onto the survey instrument. The study assumes that the MAP data collected from the Northwest Evaluation Association (NWEA) partner database were valid, reliable, and accurately reported for each school in the sample. Commonality among the participants in the sample is assumed since they are part of the same district mandate to participate in the MAP assessment.

The Role of the Researcher

The researcher is employed as a mathematics facilitator in an elementary school in the same district as the study sample population. Prior to this position, the researcher was employed as a Response to Instruction Coach within the district. Due to the various positions of the researcher in the district, participants may have known the researcher directly or indirectly. Survey participants were not identified with their responses to the researcher; however, focus group participants were able to see the researcher observe the focus group sessions. Though the researcher used a third party as a facilitator, the researcher did observe the group. This may have caused participants to feel less comfortable speaking directly about their own feelings. The questions that were asked of the focus group in the protocol were structured so that participants were asked about trends and themes, not specific questions that would require participants to reveal personal beliefs unless they chose to do so. In order to minimize researcher bias, the elementary school where she is employed was not included in the study sample.

Definition of Important Terms

Measures of academic progress (MAP). A collection of computerized adaptive assessments (NWEA, 2012).

Northwest Evaluation Association (NWEA). NWEA is a global not-for-profit educational services organization known for our flagship interim assessment, MAP® (NWEA, 2012).

Rasch UnIT (RIT, RIT score). An equal-interval measurement scale developed by NWEA and used to represent student achievement and growth on the MAP assessment. This score is used to develop national normative tables in order to measure student proficiency on each MAP assessment at each grade level. This information is also used to develop growth goals for students (NWEA, 2011b, 2012).

Proficiency. This term is used to describe the level of a student's content knowledge on a specific domain and content area compared to a national normative sample as measured by MAP as at or above the mean score (50th percentile). This is the operationalized definition of the researcher based on the results of the NWEA North Carolina Linking Study (NWEA, 2014).

Growth. A statistic appearing on some MAP reports. The growth index indicates the RIT value by which the student exceeded the projected RIT (plus values), fell short of the projected RIT (minus values), or exactly met the projected RIT (0) (NWEA, 2012).

Instructional practices. Teaching practices that engage students in the curriculum (examples include lesson planning, assessment development, instructional delivery methods, learning experiences, and environmental and materials selections made by teachers) (National Center on Accessible Instructional Materials, n.d.).

Data-based decision making (data-driven decision making [DDDM]). In education, refers to teachers systematically collecting and analyzing various types of data including input, process, outcome, and satisfaction data to guide a range of decisions to help improve the success of students and schools (Marsh & Robyn, 2006).

Summary

Research indicates that summative student performance is impacted when teachers have ongoing data about student achievement to guide their practices (Black & Wiliam, 1998). Exploring possible existing relationships between teacher capacity and attitudes associated with one CAT in comparison with student performance can serve as a guide for the allocation of resources. This study sought to illuminate associations that can impact teacher training, school-based systems and practices, and overall use of district resources to maximize student performance.

Chapter 2: Review of Literature

Introduction

This study examined how constructs within efficacy are related to student performance. This study implies that teacher capacity to use formative data from the MAP assessment and their personal feelings and attitudes towards the data and their usage have a specific relationship to student outcome measures. In order to answer the questions posed within this study, it is necessary to understand how efficacy and TE are related to DDDM and formative assessment practices. A review of the literature surrounding Bandura's (1986) social cognitive theory (SCT) as a basis for the notion of how efficacious behaviors of teachers impact student performance must be conducted. The postpositivist framework utilizes specific visual models associated with student outcomes and assessment practices that are explained with detail within this section. Specifically, two outcome models—Value-Added Assessment Model (Balls et al., 2011) and the Educational Decision-Making Loop (Masters, 2013) are reviewed as constructs that support the practice of using data from ongoing assessment practices as well as constructs related to capacity and efficacy to achieve a desired outcome.

It is also necessary to review the literature surrounding formative, interim, and summative assessment practices as they relate to DDDM. Including a review of literature surrounding CATs and how MAP assessments are currently used in educational settings is also required. A working understanding of MAP assessments and scoring is also necessary in order to answer the research questions. Each section of the literature is reviewed and then connected to this study.

Efficacy

Armor et al. (1976) from the RAND Corporation published a study entitled

Analysis of the School Preferred Reading Programs in Selected Los Angeles Minority Schools. As part of the study, a survey was sent to teachers containing two specific items that asked teachers to rate their beliefs about their own ability to bring about positive desired outcomes in student learning and engagement. Upon analysis, these items were found to measure a previously unknown construct that came to be known as efficacy (Armor et al., 1976).

Social cognitive theory (SCT). The results of studies like Armor et al. (1976) and Bandura's (1977) study coining the term observational learning theory were the foundations to what would eventually be known as SCT (Bandura, 1986). Bandura (1986) used his model of understanding how people learned through observations of situations and expanded his work to include goal-setting, self-efficacy, and self-regulation. A new focus on the psychology of learning and the cognitive processes that spurred learning led to the development of what is now SCT (Bandura, 1986). Continued expansions and evolutions of this work have continued with a renewed focus on efficacy as a major function of this model (Zimmerman & Schunk, 2001).

A link between SCT and efficacy. A major assumption within SCT is that people have an agency or ability to influence their own behavior and the environment in a purposeful, goal-directed fashion (Bandura, 2001). In this later work, Bandura (2001), written almost 25 years after his original publication that coined the theory, extended SCT to include an element of human agency that can tie together self-efficacy and the capacity to impact our environments.

TE

As the construct of efficacy emerged as a part of Bandura's (1977) SCT, research began about how this concept related to the education sector. Ashton, Olejnik, Crocker

and McAuliffe (1982) coined the term "teacher efficacy" in a paper presented at the American Education Research Association Annual meeting. Ashton refined her definition of TE throughout her research as a situation-specific expectation held by teachers that they can impact student learning (Ashton, 1983, 1984). In her 1984 study, Ashton analyzed the results of the Thematic Apperception Test (TAT) given to teachers. Three distinct discrepancies between highly efficacious teachers and teachers who report low efficacy beliefs emerged: personal accomplishment, personal responsibility for student learning, and positive belief in student performance (and behavior) (Ashton, 1984). Ashton (1984) concluded that teachers whose responses were considered highly efficacious engaged in behaviors that were different from teachers who scored differently on the TAT. Namely, they strategized about how to bring about positive results for their students by planning learning experiences and setting goals for themselves and their students related to student achievement (Ashton, 1984). Another conclusion drawn from the results of this study involved the concept of teacher attitude and its role in teacher behaviors. Ashton (1984) indicated that teacher attitudes about efficacy have a critical impact on what activities the teacher engages in related to student learning. Teachers who report high levels of efficacy hold beliefs about student learning that motivate them to make decisions related to instructional planning, time spent with students, and choosing learning experiences that have a positive impact on student achievement (Ashton & Webb, 1986). This idea led Ashton and Webb (1986) to conclude that TE consists of two dimensions—TE and personal TE (PTE).

Dimensions of TE. As the concept of TE was researched more thoroughly, expansions and revisions of the construct led researchers to conclude that TE was a multi-faceted construct (Ashton & Webb, 1986; Guskey & Passaro, 1994; Soodak & Podell,

1996). Researchers began to look more closely at Bandura's (1977) foundational work and the differences he claimed between efficacy and OE.

Guskey (1982) expanded his work on causal attribution which found that two specific attribution categories were related to teacher PE—effort and task difficulty by exploring Ashton and Webb's (1986) claims that the two-dimensional construct of TE was comprised of PTE and TE. Guskey and Passaro's (1994) study confirmed that TE was multi-dimensional; however, their results led him to categorize the distinct components as internal and external. Their conclusions were that differences in the facets of TE rooted in locus of control—what the teacher perceived he/she could and could not influence related to student performance (Guskey & Passaro, 1994).

Research continued to investigate the dimensions of TE, with a goal of separating the specific components of the construct and its relationship to OE (Gibson & Dembo, 1984; Guskey, 1982; Guskey & Passaro, 1994; Woolfolk & Hoy, 1990). Soodak and Podell (1996) conducted a study to explore the dimensions of TE to determine if there were correlations among the components and which factors within the construct related the most to impacting student learning. The study took place in urban and suburban New York where 310 teachers who represented the overall demography of the state were surveyed using a modified version of Gibson and Dembo's (1984) TE scale (TES). Analysis of the data collected led Soodak and Podell to identify three distinct factors that impact overall TE (a construct that previous research had considered a singular concept). PE was defined as the belief of a teacher that they possess the necessary skills to teach; TE was defined as the belief that teaching can overcome outside influences; and OE was defined as the belief that putting teaching skills into practice would bring about positive results in student achievement (Soodak & Podell, 1996). The major finding of this study

differentiated between PE and OE which previous research had included as one construct (Gibson & Dembo, 1984; Woolfolk & Hoy, 1990). Soodak and Podell found that the three dimensions of TE identified in their study were uncorrelated, meaning that a teacher could believe he/she has the skills necessary to teach (PE) but lack the belief that his/her instruction will impact student learning (OE). The researchers call for additional research into the construct of efficacy as a multi-faceted construct to confirm that the factors can interact in differing ways. This research is needed to inform educational leaders of how best to enhance the efficacy of teachers by identifying which dimension of TE may be lacking (Soodak & Podell, 1996).

Labone (2004) wrote about TE from the perspective of the need to expand the construct. She indicated that new research is needed to identify the aspects of TE that have the most educational impact (Labone, 2004). In essence, as research continues on both theories that inform TE and how TE impacts educational issues, more complex characteristics within the construct continue to arise (Guskey, 1982; Labone, 2004; Soodak & Podell, 1996). This study investigates an alternate paradigm within the construct of TE.

Measures of TE. Efficacy was first measured by Amor et al. (1976) when the Rand Company conducted a large-scale survey. Those items evolved into a construct known as efficacy and encompassed by Bandura (1977) in his SCT. As researchers defined a related concept known as TE (Ashton, 1983; Ashton et al., 1982; Gibson & Dembo, 1984; Guskey, 1982), the need for measurement scales emerged. Relative to this study, two primary rating scales similar to the one developed for this study have been used most frequently in TE research.

Gibson and Dembo (1984) developed the TE Survey (TES) as a part of their

study. The 30-item survey was developed to examine the relationship between teacher behaviors and TE. The instrument samples the four main domains associated with teacher effectiveness—efficacy, alignment, inclusivity, and organization (Gibson & Dembo, 1984). The TES was administered to 55 teachers and data analysis unexpectedly yielded two distinct factors. The researchers termed these factors PTE and TE. The researchers related these factors to Bandura's (1976) social learning theory stating that PTE was related to self-efficacy and TE was related to OE (Gibson & Dembo, 1984). The interpretation of these factors has been debated by researchers (Soodak & Podell, 1996; Tschannen-Moran et al., 1998; Woolfolk & Hoy, 1990).

Citing concerns with the construct validity and statistical instability of the TES, Tschannen-Moran et al. (1998) analyzed the intercorrelations of all available TE measures in their study. As a part of their study, Tschannen-Moran et al. developed a new measure of TE known at the time as the Ohio State TES (OSTES) and is now referred to as the Teacher Sense of Efficacy Scale (TSES). The scale is a 24-item survey that uses a 9-point scale with five descriptor markers ranging from "none at all" to "a great deal," describing the level of influence a teacher believes he/she has related to the item. Three primary factors were measured on the scale: Instruction, Engagement, and Management (Tschannen-Moran et al., 1998). The researchers concluded that their new measure advanced the field of TESs because of the "unified and stable factor structures" that were missing from previous measures (Tschannen-Moran et al., 1998, p. 801).

Models of OE

The cycle of teacher efficacy judgments. Tschannen-Moran et al. (1998) released an idea about how teacher efficacy beliefs relate to student outcomes. This cyclical graphic of teacher efficacy judgments is a model of their findings.

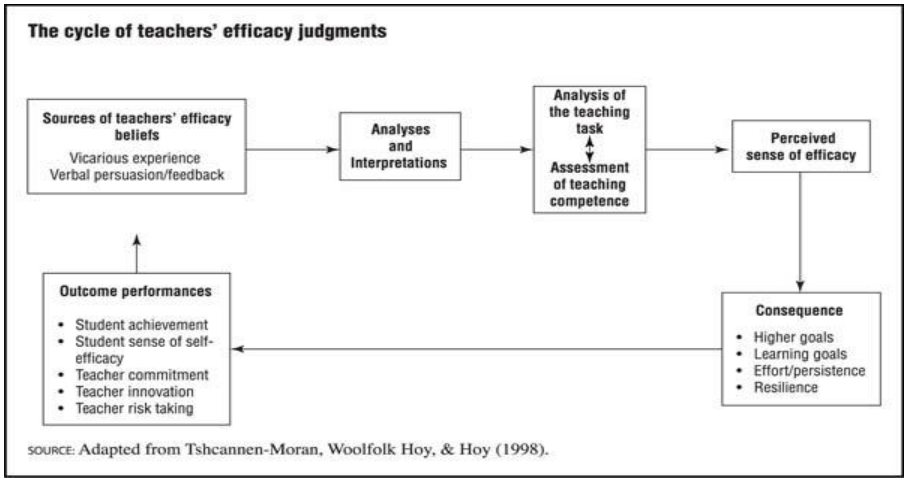


Figure 1. *The Cycle of Teacher Efficacy Judgments.*

Tschannen-Moran et al. (1998) reviewed TE studies and found that the construct was multi-faceted and that measures lacked a common definition of the paradigm. With the development of the TSES, Tschannen-Moran and Woolfolk Hoy (2001) formulated an operational definition of TE as "a teacher's perception of their resources and strategies for bringing about student behavioral and instructional outcomes shift" (p. 784). The minor changes in the questions on this self-report scale shifted the concept of TE as a rating of the overall confidence of the teacher to a mindset of teacher beliefs about their capacity to impact the achievement outcomes of their students. The new TSES instrument entwined TE with their self-reported capacity to bring about the results associated with student achievement (Tschannen-Moran & Woolfolk Hoy, 2001).

Value-Added Assessment Model. Balls et al. (2011) reported the increase of focus on value-added models of student achievement as a technique used to measure longitudinal growth. Of the dual focus for these models identified by Balls et al., the predictive component of student achievement on high-stakes tests supports the need for efficacious teaching practices surrounding formative assessment (Black & Wiliam,

2009). In the text, the authors state that the level of skill of the teacher to facilitate learning is the key to student progress and achievement; but while this fact is acknowledged, the purpose of the value-added model involves changing the learning culture of a building to include reflective components of teacher self-efficacy and collective efficacy developed through shared decision making. It is the conclusion of this model that improvements to these structures within a learning system will lead to increased student outcomes (Balls et al., 2011).

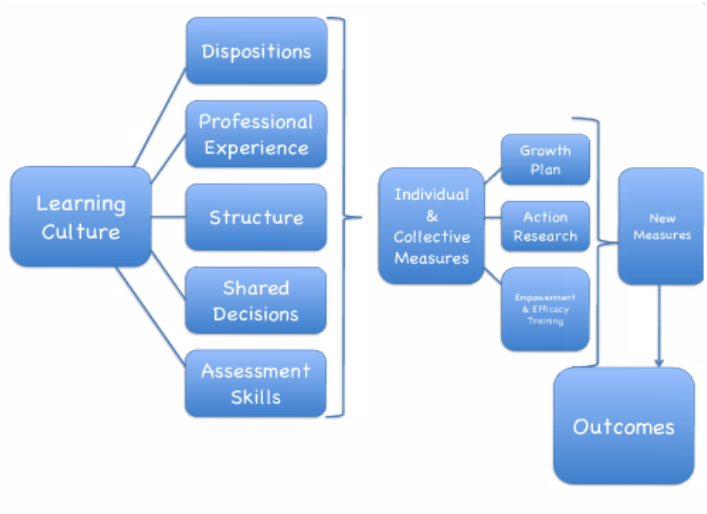


Figure 2. *Value-Added Assessment Model.*

Educational Decision-Making Loop. Masters (2013) illustrated his Educational Decision-Making Loop (p. 10) as a representation of how feedback can be utilized to inform future practice. This decision-making process derived from the increase in demand for improving student outcomes on assessment measures (Black & Wiliam, 1998). Masters claimed that professional action is central to the process of decision making and that action should be driven by considerations of current situations and prior knowledge about previous outcomes. This model is useful in all elements of educational action planning and assessment because it can be utilized by teachers considering what

their students already know about a topic and what is needed to improve student outcomes. The recursive nature of the model allows for the process to continue as an ongoing part of instructional assessment and decision making (Masters, 2013). The same process can occur at all levels of an educational organization where the main goal is to improve student outcomes. This model is representative of any data collection and analysis model of continuous improvement that has a goal of improving life consequences for students (Masters, 2013).

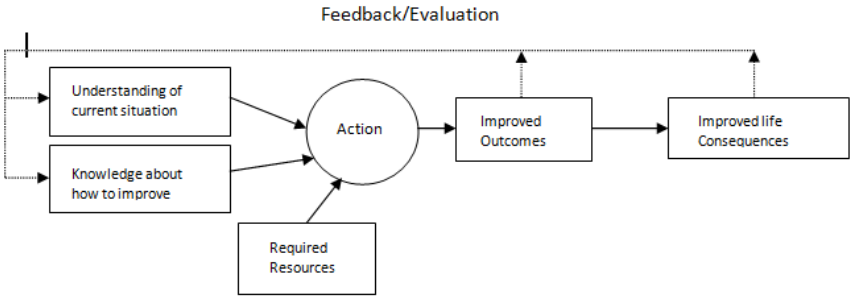


Figure 3. *Educational Decision-Making Loop.*

Like the value-added model introduced by Balls et al. (2011), Masters' (2013) Educational Decision-Making Loop can be used as a way to engage in continuous cycles of improvement by involving educators as action researchers, using data and expert knowledge to evaluate actions and processes that lead to the improvement of student achievement outcomes. More apparent in the value-added model, the efficacy of teachers plays a role in both improvement models. The value-added model stresses the need for teachers to believe they can improve outcomes for their students, whereas the Educational Decision-Making Loop focuses on what teachers know about the areas of improvement necessary for increases in achievement outcomes (Balls et al., 2011;

Masters, 2013). Positive associations between teacher capacity and attitude related to student performance would confirm the use of representative models such as these as part of educational best practice.

DDDM

Similar to OE models for understanding how TE and student performance are related and how balanced data systems integrate all forms of assessment, graphic representations are used in education to represent ways to interpret and use data to impact instructional practices. As earlier defined, DDDM refers to teachers systematically collecting and analyzing various types of data including input, process, outcome, and satisfaction data to guide a range of decisions to help improve the success of students and schools (Marsh & Robyn, 2006).

The USDOE published "Implementing Data-Informed Decision Making in Schools: Teacher Access, Supports and Use" outlining best-practice recommendations for school districts initiating a data-driven instructional model for student learning (Means, Padilla, DeBarger, & Bakia, 2009). Specific guidelines included (1) curriculum-aligned benchmark assessments; (2) teacher buy-in to the data system and continual maintenance of teacher perceptions; (3) professional development for interpretation of data and translation of data into instructional practices; (4) removal of policies that serve to prohibit the use of data by teachers to inform instruction; and (5) school leaders should create an environment of mutual trust among colleagues for using data to reflect on teaching practices (Means et al., 2009, pp. 63-65).

Even before the guidelines from the USDOE, Boudett, City, and Murnane (2006), in conjunction with school districts, created the Data Wise Improvement Model which is an 8-step cyclical graphic of the process schools can use to manage the data flow,

decipher the information, and make instructional decisions. The function of the model is to improve student learning using available assessment data by creating a process to examine all the data in a manageable way that leads to positive, measurable results (Boudett et al., 2006). The Data Wise Improvement Model is broken into eight steps over three stages—Prepare, Inquire, Act. Boudett et al.'s model is designed for school personnel to cycle through continuously, as needed, until the desired result is achieved akin to current continuous improvement models trending in both the corporate and social sectors as pathways to success.

This model indicates that the two initial steps in this process are foundational, involving creating an inquiry-based culture and designation time and protocols for data discussion and educating the staff in data literacy so they have the necessary ability to use the data obtained from assessments. This model supports the need for teacher capacity with data practices to be a focal point of educational training (Boudett et al., 2006). Steps 3-5 of the inquiry phase of the model focus on the examination and integration of available data, alongside an examination of instruction from planning to delivery before moving into the action planning stage. Developing an action plan involves building a guide for implementation and assessment to determine effectiveness of the plan (Boudett et al., 2006, p. 3). The Data Wise model requires that schools continually evaluate their plans, their effectiveness, and the student learning impacted to determine next steps. This process for a school where new data comes in each day functions in a constant loop for improving student learning outcomes (Boudett et al., 2006).

An urban perspective. Heppen et al. (2011), along with the Council of the Great City Schools, published an urban data study entitled Using Data to Improve Instruction in the Great City Schools: Documenting Current Practice, examining the current data

practices surrounding interim assessments in four districts. The study consisted of survey data collected from curriculum coordinators (CCs) and research directors (RDs) and in-depth site visits investigating the use and practice of data obtained from interim assessments in their districts. This project, funded by the Bill and Melinda Gates Foundation from 2009-2011, had two goals—to examine the current data practices in urban schools districts related specifically to the administration and use of interim assessments and data and to determine a relationship between student achievement and school/classroom data practices. This study sought to set forth best practices in the use of interim assessment data to improve instructional practices and guide educational decision making. In the initial phase of this study, researchers surveyed the CCs and RDs in the Great City Schools asking about data usage, interim assessment practices, and data systems using surveys tailored to each specific job title. The survey, opened from January to September 2009, yielded responses from 94% of the districts in the sample, meaning that either the CC, RD, or both responded to the survey. Results of the survey data indicated that there were data systems in place to create, administer, and use data from interim assessments in an overwhelming number of districts in the sample (98%). The survey data indicated that typically interim assessments were given three times a year (MAP benchmarks are given in this manner) and as many as seven times a year and typically in reading (ALL districts reporting) and mathematics (94%). The study found that interim assessments were used in Grades K-12, with the highest reported use (80%) used across the reporting districts in Grades 3-8. It was found that 75% or higher of CCs and RDs reported that interim assessment data were used for the following: Guide and Inform Instruction; Formative; Diagnostic; and Measure Progress towards End-of-Year assessments. Asked to only CCs, 15% reported that the data from these assessments were

used for rewards or sanctions. Eighty-eight percent of CCs responded strongly agree or agree that their district had allocated substantial resources to the use of data to guide instructional decisions; 70% reported that there were expectations in place for the use of interim assessment data in their districts; while 82% agreed that their district had articulated clear goals for the use of assessment data. Of specific note to this study is that 85% of CCs reported that the expectations from district leaders were that interim assessment data should be used by principals, building leaders, and teachers to inform instructional decisions (Heppen et al., 2011).

The second phase of the study involved an in-depth case study of four districts in the Great City Schools who met inclusion criteria set by the researchers. The 2-day data collection involved schools of various sizes in differing geographic areas. Based on survey results of interim assessment administration rates for third through eighth grades, the qualitative portion of the study focused specifically on Grades 4, 5, 7, and 8. Focus groups of four to 10 participants were assembled randomly from district staff, teachers, and principals. A total of 56 teachers, 40 district-level staff, and 28 principals were included in the sample size from the four districts. The mean of the four districts' student populations approximates 130,750 students (the district in this study has a similar approximate student population of 145,000). Five themes emerged regarding goals for interim assessment usage: increase classroom instructional accountability, ensure consistent monitoring of school/student progress, use by teachers as a tool to guide classroom instructional practices, prepare students for/predict student performance on state assessments, and inform school improvement planning. A common challenge reported in all four districts was the communication of the expectations regarding the use of interim assessments that centered on a lack of teacher understanding of the purpose or

how or why to use the data. Heppen et al. (2011) found that teachers reported giving the assessments because of a mandate but not really using the information as a way of assessing student performance, defaulting to teacher-made assessments or classroom data. Teachers reported that sometimes they felt they were being evaluated by the assessments, even though survey results indicated that CCs only reported that to be the case in 15% of the districts. Among the four districts, data infrastructures for data storage, retrieval, and support were similar. Differences were noted from the focus group and interview data in the manner and frequency of data-based conversations and planning. Data practices varied from school to school and were reported to be possibly related to principal and leadership buy-in and ability to use the information provided by the assessments. This qualitative data indicated that there is a self-reported association between capacity to use and attitude towards the use of formative data from interim assessments and creating a culture of data usage in schools and districts. Most of the teachers reported that training provided by their districts did not include how to use the student data from the interim assessments to change their classroom planning and instruction. This theme was supported by specific quotes from teachers who indicated that any successes they were having with data usage came as a result of discussing what was going on in each teacher's classroom and for specific students. This level of analysis was necessary for the data to be used to inform instructional practices. Five themes emerged for current ways interim data were used to inform instructional practices: differentiated instruction, remediation/reteaching, grouping of students, identifying students for tutoring, and goal setting with students. The researchers noted that despite differences in districts on various elements of the study, one commonality was that resources are being directed towards interim assessment and the use of the data to make

classroom, school, and district educational decisions. This portion of the longitudinal study served to illuminate current data practices. Further statistical analysis continues in the next phase of the research to determine best practices associated with using interim assessment data to make instructional decisions (Heppen et al., 2011).

Assessment

Summative assessment. This form of assessment examines levels of student proficiency at a culminating stage, such as an end-of-grade (EOG) state assessment (Henderson, Petrosino, Guckenburg, & Hamilton, 2007). This type of assessment increased as an important practice following the 2001 NCLB Act that called attention to measuring student performance from year to year. Recent RttT initiatives have tied performance on summative assessments to funding levels and teacher performance evaluations, focusing efforts of educational leaders to find ways to made better predictions about student summative performance on these assessments (Pinchok & Brandt, 2009). Henderson et al. (2007) stated that this assessment practice is "designed to show the extent to which students understand the skills, objectives, and content of a program of study" (p. 2)—or mastery learning assessment.

Perie, Marion, and Gong (2007) published a study examining various assessment practices and their potential implications on student learning. In the study, the researchers devoted time to defining different assessment practices and integrated those definitions into current ways that assessments are presently being categorized and used by schools (Perie et al., 2007). According to that of researchers, summative assessments are designed to be given at the end of a cycle (i.e., grade; unit of study) and are not designed to provide ongoing instructional feedback to educators in order for adjustments to teaching practices that would impact student achievement in that cycle. The

researchers cited this as a major flaw when summative assessments are used in isolation and noted that this fact has spurred the practices of more frequent assessment of student learning prior to summative components (Perie et al., 2007).

Formative assessment. Black and Wiliam (2009) concluded that when classroom assessments are ongoing and their data are used for planning, teachers are more likely to make more instructional-sound decisions. The practice described by Black and Wiliam is known in education as formative assessment. According to the Center on Response to Intervention, formative assessment is a "form of evaluation used to plan instruction in a recursive way" (Center for Response to Intervention, 2014, p. 4). Characteristics of this evaluation method include the regular assessment of student knowledge and skill, growth measures, and diagnosis of areas of concern (Fuchs & Fuchs, 2005). The practice centers on the use of formative information about student performance to guide teachers in instructional DDDM. Having information in this way in order to plan next steps in instruction for a student is shown to lead to better decisions than when those data are absent (Black & Wiliam, 2009). Pinchock and Brandt (2009) asserted that engaging in research-based formative assessment practices will lead to better instructional practices in the classroom. In the conclusion of their 2009 article, Black and Wiliam stated,

> Thus, whilst we cannot argue that development of formative assessment is the only way, or even the best way, to open up a broader range of desirable changes in classroom learning, we can see that it may be peculiarly effective, in part because the quality of interactive feedback is a critical feature in determining the quality of learning activity, and is therefore a central feature of pedagogy. (p. 100)

Formative and summative association. Weiner and Hall (2004) found that central to the NCLB Act is the accountability of school districts and states measured by adequate yearly progress (AYP). The basic premise involves setting goals for student learning, measuring student progress, and ensuring that goals set forth for the end of the year are met; as measured by the state's summative assessment. This policy makes it necessary for schools to have a clear understanding of which students are meeting learning targets and which are not. Initially, this was done only with summative examinations, but as accountability increased under governmental funding and legislation, the need for measuring student learning in an ongoing method was highlighted. This allowed for schools and districts to make instructional decisions in the moment that could positively impact the trajectory of student learning. AYP was designed to be a screener to determine if specific schools were meeting the needs of their students. Weiner and Hall conducted case studies of several schools in the U.S. that have successfully met or exceeded the AYP requirements. One of the many key findings was that common to the schools was frequent diagnostic assessment. The information captured from these assessments has been used in these successful schools to make instructional decisions about student learning needs. Teachers in these schools use the information to guide their lesson planning and delivery for students based on their strengths and needs (Weiner & Hall, 2004).

Interim assessment. Another assessment practice that combines mastery assessment with ongoing specific feedback (Perie et al., 2007) that can be used for instructional DDDM (Henderson et al., 2007) is called interim assessment. Interim assessments are curriculum-aligned and designed to be given multiple times during a school year and provide timely feedback that can be used to determine what instruction is

necessary for specific students; illuminate trends in classrooms, grade levels, and schools; and contains a predictive element that measures student growth across time (Henderson et al., 2007; Perie et al., 2007; Pinchok & Brandt, 2009).

Since interim assessments integrate components of both formative and summative practices, they can be used in both ways (Henderson et al., 2007). Often referred to as benchmark testing, these assessments are formative in that they provide information about student levels of proficiency on specific skills related to the curriculum in a timely manner that allows for instructional decision making using the data obtained from student performance on the measure (Henderson et al., 2007). In addition, because interim assessments can be given multiple times across a year, they can be used as growth measures; a more summative approach to the data (Henderson et al., 2007; Perie et al., 2007). The predictive element of interim assessment has been popularized by developers of these assessments as a way to determine a student's likelihood of scoring proficiently on state assessments, tapping into the summative component of this assessment type (Perie et al., 2007).

Interim summative proficiency comparison. With regulations set forth by NCLB, states were given the autonomy to determine proficiency levels for their specific state without any input from the federal government (Cronin, Dahlin, Adkins, Kingsbury, 2007). This has resulted in a problematic definition of the term "proficiency" from state to state and even assessment to assessment. In "The Proficiency Illusion," Bracey (2007) examined definitions of the term from various credited assessment programs such as NAEP and NWEA. Since NAEP is a nationally recognized institution in assessment in the U.S., Bracey referenced that NAEP's definition is often accepted as to the "gold-standard" (p. 316). Bracey cautioned that NAEP is not an assessment that is aligned to a

specific state curriculum, so that as long as states in the U.S. are operating under independent curriculums, this would not be an appropriate strategy for defining proficiency. Bracey reported that MAP assessments have the ability to use an item bank that is aligned to state standards, along with an equal-interval scale that allows NWEA researchers to compare student MAP scores to state assessment scores to create a "linking study" where it is possible to see associations between MAP scores and summative performance scores (p. 316). It was discovered that though this provides helpful information at the specific state level, because "cut scores" needed to pass state assessment vary tremendously, a specific MAP proficiency level cannot be ascertained at the national level (Bracey, 2007, p. 317). Bracey cited the example of a student in Texas who would only need to be in the 12th percentile according to MAP to be considered proficient on her summative state assessment, where a student in California would need to be in the 61st percentile to be considered proficient on the California state assessment. Discrepancies like these continue to make a finite definition of proficiency illusive, requiring states and districts to operate independently to set their own criteria for proficiency (Bracey, 2007). Bracey (2007) referenced NWEA's own published findings that cut scores for proficiency are varied from mathematics to reading as well as among grade levels. Specifically, it is stated that educators should be cautioned that the whole construct of proficiency of MAP assessment scores is an illusion because students whose data indicated that they were on trajectories for successful proficiency in lower grades can be derailed by just the arbitrary change to the cut score at a higher grade. Bracey summarized the cautions of his paper by indicating that those who claim success of NCLB legislation in increasing the number of U.S. students proficient in math and reading should consider that this may not be the case at all but more a result of the

increasing ease of the assessments used to measure proficiency. This indicates that the real information gained from MAP assessments lies in the power to understand which concepts were strengths and which were weaknesses for students so teachers could use the data to guide instructional decisions that impact learning. If states and districts are only using the assessments to meet the requirements of federal legislation associated with NCLB (Bracey, 2007), they are missing the part of the assessment that is valuable to instructional practices independent of performance on summative assessments or proficiency cut scores.

Balanced Data Systems

In their article, Perie et al. (2007) coined the Tiers of Assessment model which is used to show how the three assessment categories can be integrated to form a balanced data system.

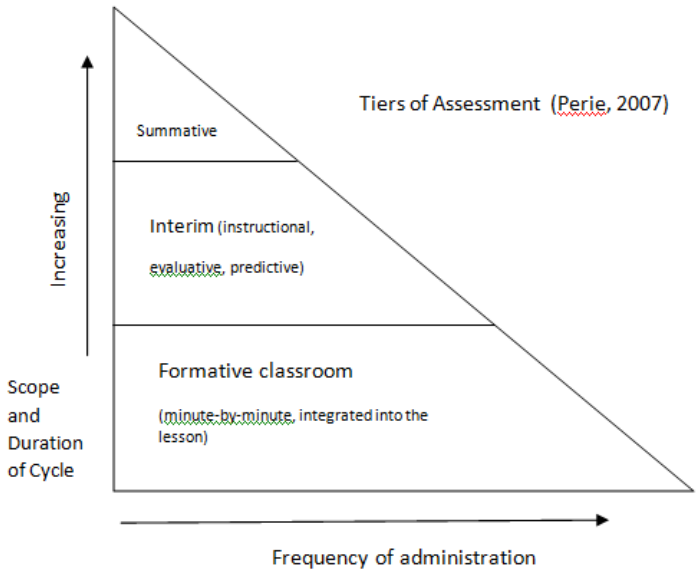


Figure 4. *Tiers of Assessment.*

This model shows how formative assessment is the most frequent type of

assessment with the shortest duration and smallest scope of the curriculum, while summative assessment is the least frequent but takes the most time and encompasses the most curriculum components. Interim assessment is a median between the two more extreme forms (Perie et al., 2007). Perie et al. (2007) magnified that data systems that were successful at impacting student achievement utilized each of these assessment types judiciously depending on what information they needed from the assessment to make instructional decisions. A review of the literature on assessment supports that educational institutions with a data-literate culture are balanced in their assessment practices (Black & Wiliam, 2009; Henderson et al., 2007; Olson, 2001; Perie et al., 2007; Pinchok & Brandt, 2009).

CAT. According to an article published by the Kingsbury Center, which is the research arm of NWEA (developers of MAP CAT), Wang et al. (2013) defined CAT as an achievement test where students are presented with many different items that have been tailored to their ability levels. These items are adjusted in real time during the assessment based on the students' responses to previous items (Wang et al., 2013).

In the 1960s, research began in the Office of Naval Research (ONR) investigating ways for examinees to respond to various questions of similar content and construct that were adaptive to their specific ability levels. As technology advancements emerged, the Department of Defense (DoD) sought to apply their findings about adaptive assessment to include computer-based administration. Though research and design continued for more than 20 years, it was not until 1991when the U.S. military researchers presented to NATO a workshop on the practice of what came to be called CAT. The military cited that once technological advancements could be achieved, the practice of CAT would save "valuable resources" (Sellman, 1988, as cited in Sands, Waters, & McBride, 1997) for the

military. Specific benefits in cost benefits, time allocation, and data analysis mentioned in Sands et al.'s (1997) article are also applicable to the field of education.

Woodfield and Lewis (2003) reviewed Idaho's change from the traditional paper-pencil high-stakes testing to an online assessment. Idaho was the first state in the U.S. to boast such a change as a response to concerns raised by teachers and district leaders about the need for an assessment that measured and tracked student progress and growth. Idaho's educational leaders called for assessments that would provide data that could be used to inform instructional decision making in time for teachers to adjust classroom practices that could impact student achievement. The consensus from state leaders was to use the MAP assessment from NWEA, which was already being utilized in several Idaho districts. It was determined that MAP assessments offered solutions for the major concerns raised by educators. Woodfield and Lewis indicated that not only was the technology used by NWEA to administer and score MAP assessments feasible for easy implementation in the state, the norm-referenced assessments aligned to state standards offered the growth data that was missing in previous assessments. According to the authors, the data from the MAP assessments provided detailed information that furthered the practices of DDDM in districts across the state. Woodfield and Lewis highlighted the Rasch index (RIT) used by MAP, a measurement of student performance, and the major difference between MAP assessments and other computerized assessments. The equal-interval scale provided the means for teachers to assess student understanding on a specific concept along a fluid continuum and easily determine growth. Clark (2004) claimed that using the MAP-like systems is better than traditional assessments because the data are about student growth and proficiency, not about comparison to other students.

MAP

Definition. By definition of the developer, MAP is a "collection of computerized adaptive assessments" (NWEA, 2012, p. 4). According to one elementary school in the district where the study was conducted,

> MAP Testing will take place for all students from Kindergarten through 5th grade. MAP, or the Measure of Academic Progress [assessment], is a computerized adaptive test which helps teachers, parents and administrators improve learning for all students and make informed decisions to promote a child's academic growth. (School website, 2014)

Another elementary school in the same district described MAP as "a national norm-referenced test as a diagnostic tool to help teachers customize instruction on a timely basis" (School website, 2014). The district where the sample population was studied stated that it "uses MAP as an instructional tool to help teachers instruct students exactly where they will learn best" (District website, 2014). It is important to this study to include local descriptions of MAP assessments because they speak to the perceptions of and general context of usage of these assessments within the population this study seeks to inform.

As essential as it is to have a working definition of the MAP assessment, understanding the ways in which these assessments are used in educational practices are even more crucial to this study.

Uses for MAP. According to NWEA (2014),

> Whether you need an interim test to benchmark student growth, or a universal screener for early learners, MAP can help. Administered online or locally, MAP helps you pinpoint—to the goal—strand level—where your students are ready to

> advance, and where they need help. And thanks to our stable, empirically-derived RIT scale, longitudinal data from MAP can be used to analyze program impact—a huge help in challenging budgetary times. (MAP Overview home screen).

This is taken from the marketing section of the website under the "Solutions" tab. Since uses for MAP can vary among Local Education Agencies (LEAs), it is necessary to explain the use of MAP in the district from which the participants were selected.

Wang et al. (2013) conducted a longitudinal study to examine student achievement across time and grade levels as measured by CAT. The study focused on the internal validity and invariance of a CAT achievement test. The scope of the study was to examine the constructs on a large-scale CAT assessment in 10 states. Between spring 2009 and spring 2011, MAP data were collected from 10 states with the largest samplings of available data. Though the assessment was administered to students in Grades 3-10, the sample in this study began with students in fifth grade, continuing with that sample through seventh grade. The study used an analysis of factorial invariance and a Multiple-Indicator Latent Growth Model (MLGM) and concluded that their findings supported that MAP interpretations were "consistent and reliable" (Wang et al., 2013, p. 43) across the states in the sample. The researchers concluded that MAP was a CAT that had internal validity and invariance with regards to student achievement that was consistent across the longitudinal study (Wang et al., 2013).

Universal screener. In the population school district from which the study sample was obtained, MAP is used in two major ways. The initial intent of the program in the district was as a universal screening measure as a part of the Responsiveness to Instruction (RtI) framework that began in the district in the 2012-2013 school year (personal communication, 2013).

As a component of an RtI framework, universal screening is the first step in identifying the students who may be at risk for learning difficulties. Universal screening is typically conducted as a benchmark assessment three times per school year in the fall, winter, and spring. According to Jenkins (2003), universal screening measures consist of "brief assessments . . . that are highly predictive of future outcomes" (Jenkins, 2003, as cited in Hughes & Dexter, 2011, p. 1).

The population district continues to employ MAP as a universal screener, but since these formative benchmark data are now available for every student in kindergarten through eighth grade, the information is used as a part of instructional decision making throughout the district by teachers, administrators, and executive personnel. Ongoing training is provided at the district level to support the use of MAP data to inform instructional practices. Though individual schools use MAP data in various ways and different levels of degree, it is basic assumption of the school district that the data are utilized as a part of a data-based continuous improvement model (personal communication, 2014).

Universal screening practices should be a normative-referenced assessment given to all students, using scores at or below the 25th percentile as an indicator of students who may be academically at-risk (Fuchs et al., 2007). This screening practice in an RtI framework provides teachers access to performance data about all students which allows for instructional decision making to extend beyond the design of an identification process for at-risk students (Jenkins, 2003).

Growth and proficiency measure. The benefit of the design of the RIT is that it allows results on the MAP assessment to be compared among states on an equal measurement scale. Cut scores often used to determine levels of proficiency could then

be stabilized across all sates using MAP as a measure of student achievement. This addresses an issue posed in the Bracey (2007) report, "Are proficiency levels consistent and comparable among states" [or teachers] (p. 317)? This question is at the root of NCLB funding because the law calls for states to set in place their own proficiency criteria; however, funding levels are tied to student proficiency (NCLB, 2002). This study highlights the benefits in having a measure from which a common proficiency level can be determined among states; therefore, equalizing accessibility to funding requirements from the federal government (Kingsbury, Olson, Cronin, Hauser, & Houser, 2003).

Kingsbury et al. (2003) conducted a meta-analysis of 14 other research studies conducted between 1997 and 2003 that examined the proficiency standards established in various states. The study concluded that proficiency levels varied significantly among states ranging in some grade levels and content areas from 67% to 13% as proficiency "cut scores." The study also found that consistency varied within states in regards to between-grade and between-content constructs. This study illuminates the need for assessments with internally valid scores. According to Wang et al. (2013), MAP assessment results are considered valid and reliable. Since research indicates strength in correlation at the national level (Cronin & Bowe, 2005; Kingsbury et al., 2003) and the state level where the current study was conducted (NWEA Linking Study) between student achievement on MAP interim assessments and summative performance on the state EOG test, resources will likely continue to be allocated for MAP CAT assessments within the population district.

MAP alignment research. In 2005, Cronin and Bowe (2005) published a linking study that examined the alignment of NWEA MAP scores to the Arizona end-of-year

summative assessment (AIMS). NWEA conducts these linking studies in all states where MAP is used as formative data to provide states information about how those data compare to student proficiency on their specific state summative assessment. The purpose is to align specific RIT scores from MAP with specific proficiency levels on the AIMS assessment to establish "cut-scores." The research also examined how well student performance on AIMS can be predicted by MAP given in the prior spring and fall administrations. The study analyzed data from 15,000 students in three Arizona school districts, excluding those receiving accommodations on their AIMS, in Grades 3-8 on both reading and math assessments. Strong positive correlations (between .79 and .85 in reading and .84 and .88 in mathematics) were found between MAP scores and AIMS performance. Correlation was the strongest between student scores on the spring administration of the MAP assessment and the AIMS. It was believed that this was due to the close proximity of the spring MAP test administration and the AIMS administration. In reading, the strength of the correlation decreased as grade levels increased but not with any statistical significance. In math, the correlation remains stable and strong between third and eighth grades. The researchers created a cut-score table comparison, matching RIT scores (MAP) to the criteria from the AIMS assessment of Exceeds, Meets, Approaches, and Far Below. A predictive index was calculated and determined to be between .87 and .90 for using RIT scores from MAP to predict proficiency levels on the AIMS. The study found that though the predictive indices were high for math and reading across grade levels, the researchers caution educators that students who near the NWEA-MAP proficiency score only had about a 50% chance of passing the AIMS. Cronin and Bowe noted that this has also been found to be true of data analyzed from other states. This is believed to be because proficiency on state

assessment is not aligned to national normative data; rather, more arbitrary cut scores. The researcher concluded that students needed to show much higher proficiency levels according to MAP to consistently pass the AIMS summative assessment (Cronin & Bowe, 2005).

Cronin and Bowe's (2005) findings of stronger correlation between the data collected from the spring administration of MAP and the summative AIMS assessment informed the decision of the researcher to use spring MAP data in this study. Though formative-summative correlations are beyond the scope of this study, the value of the correlation between formative data and summative performance is a central construct to the importance of data-based decisions in educational practices. Cronin and Bowe's conclusions also cautioned against the 50th percentile as a predictive indicator for summative proficiency; however, for the purpose of this study, the researcher relied on data from the linking study for the state of the participant sample to guide the decision to utilize the MAP normative mean (50th percentile) as an operational definition of proficiency (NWEA, 2014).

In March of 2014, NWEA released an alignment study between NWEA RIT scores and the North Carolina EOG assessments. The findings were based on a sample of 18,730 North Carolina students who took both the math and reading EOGs in the spring of 2013. The NWEA researchers used an Equipercentile calculation method to estimate the RIT score equivalent to each of the five state performance levels in North Carolina. The method used the following procedure:

> We determined the percentage of the population within the selected study group that performed at each level on the state test and found the equivalent percentile ranges within the NWEA dataset to estimate the cut scores. For example, if 40%

> of the study group population in grade 3 mathematics performed below the proficient level on the state test, we would find the RIT score that would be equivalent to the 40th percentile for the study population (this would not be the same as the 40th percentile in the NWEA norms). This RIT score would be the estimated point on the NWEA RIT scale that would be equivalent to the minimum score for proficiency on the state test. (NWEA, 2014, p. 2)

From this data, NWEA researchers set cut scores which they defined as "the minimum estimated score" necessary to score a corresponding level on the North Carolina EOG assessment in either mathematics or reading (p. 3). Estimated probability calculation tables were also created comparing the entire range of RIT scores possible to the probability percentage of scoring proficient on the state assessment for each grade level in both subjects (NWEA, 2014).

Much like the Cronin and Bowe (2005) alignment study, Pearson's *r* correlation was calculated between MAP and the EOG for each grade and test subject. Strong positive correlations were seen at every grade level and for both subject areas (math 0.814-0.839 and reading 0.821-0.775). Just like in the Cronin and Bowe study in Arizona, correlation decreased, albeit not significantly, in reading as grade levels increased (NWEA, 2014).

Based on the strength of correlation between MAP and North Carolina EOG scores, the researcher determined for the purposes of this study that the information from the linking study would be used to operationalize the definition of proficiency at the 50th percentile (NWEA, 2014, Table Set 1).

MAP and TE. Bremer (2008) examined the impact of TE related to the MAP formative assessment program as an indicator of student performance on the North

Dakota state examination. This was a comparison of how teachers felt about and utilized the data from the computerized-adaptive formative MAP assessment from NWEA as a possible indicator of success on a summative assessment aligned to the same standards. The sample was derived from the population of all teachers in North Dakota who taught Grades 4-6 and any seventh- and eleventh-grade teachers who taught math or language arts. The sample included teachers from 14 of the largest school districts in North Dakota, though Bremer noted that these districts would be considered small in a state with more urban areas. The minority populations in these districts included five above the state average of 9% and one with a rate of 28%. The researcher noted that there are a low number of students in the state living in poverty.

Bremer (2008) used a purposive, nonrandom sampling to obtain her participants. Permission to conduct research was offered to all school districts in the state, though superintendent and principal permissions were required. Ultimately, 64 schools joined the study; and of those, only 10 represented districts used the NWEA MAP assessments. The researcher sent surveys to 508 teachers who had been identified by their principals as meeting the study criteria, of which 162 were returned with all four measures of efficacy completed. Of the teachers from districts participating in MAP assessments, 139 usable surveys were returned for a response rate of 31% (Bremer, 2008).

Data were collected on how and in what ways MAP score information was utilized, as well as two items that measured teachers' perceptions about helping students perform on state assessments. Bremer (2008) used the TSES (Tschannen-Moran & Woolfolk Hoy, 2001) and the Collective Teacher Beliefs (CTB) scale and then developed two independent scales that adapted each of these rating scales to specifically ask about the North Dakota State Assessment (NDSA). Bremer's assessments, the Teachers'

Beliefs about the State Assessment (TBSA) and Collective Teachers' Beliefs about the State Assessment (CTBSA), related to teachers' beliefs about their ability to influence student scores on the NDSA. The researcher used a survey development method that included writing questions, vetting them among colleagues, making revisions, piloting the survey, and using psychometrics to determine reliability and validity of questions. In order to increase sensitivity, Bremer cited Bandura's (2001) recommendations and used a 9-point scale ranging from 1—"Cannot do at all," to 9—"Highly certain can do" (p. 54). Respondents were asked to rate their perceptions of their capabilities in relation to eliciting specific behaviors from their students that would impact their performance on the NDSA. Bremer also collected teacher and school use of MAP data in forms she called the Teacher Use of score information (TUSE) and School Use of score information (SUSE). These consisted of four items that could be checked off indicating ways that teachers use the data collected from MAP assessment in instructional practices.

A causal comparative research model was used to identify factors that contribute to teacher perception of efficacy related to their ability to impact student performance on state assessments (Bremer, 2008). The researchers created two groups (MAP and non-MAP users) to compare their feelings of efficacy related to if they could make a difference in student performance on the NDSA. The researcher conducted prediction studies between measures of efficacy and assumed using MAP information "precedes the development of stronger efficacy beliefs in teachers" (Bremer, 2008, p. 58) using multiple regression with significance levels of $p<.05$. Statistical analyses of the data were conducted on the means of groups based on demography and teacher use of MAP data. Both *t* tests and ANOVA were used to examine differences in the means. Bremer (2008) followed up with regression analyses to consider predictive factors of efficacy

scores.

The results of the one-way ANOVA conducted on the relationship between teachers' responses to the efficacy scales and student performance on the NDSA indicated significant results at R^2=.06 (p<.05). Frequency distributions were created based on the 138 viable responses on the TUSE and SUSE for each possible item checked. TUSE Item 1: "I look over my students' scores on the MAP tests" had a frequency of 84%; and SUSE Item 1: "Teachers in this school look at the MAP scores of their students" had a frequency of 80% (Bremer, 2008, pp. 78-79). These items were considered the lowest level of use on the surveys. The other items fell between 42-54 percentages checked on both measures. Post hoc analysis of the data indicated that there was a significant difference in efficacy perceptions between teachers who used MAP data in at least four ways and those who used it in only one way. This supported the hypothesis made by researcher (Bremer, 2008).

Limitations of this study included the lack of ability to conduct data analysis of the relationship between teacher perception of efficacy and teacher use of MAP data due to lack of an appropriate number of responses. In addition, the lack of diverse minority populations and ability to include large, urban school districts limits generalizability of the results. Bremer (2008) noted additionally that though her two survey instruments were piloted, they had not previously been used for comparison of validity or reliability constructs. This study examined the relationship between general efficacy constructs and use of MAP data and student summative performance on a state assessment; however, there is a call for research where teacher perceptions about the data from MAP assessments and student performance is compared.

Bremer (2008) concluded that the study reflects that staff development and

training related to using MAP data to drive instructional practices is needed in conjunction with promoting the understanding of how positive efficacy in teacher beliefs impacts student performance. Final recommendations of this study include an amendment to NCLB which allows states to use adaptive assessments, like MAP, for growth indices related to measures of AYP, a measure that accounts for school rankings, funding levels, and status. Bremer cited the flexibility of the MAP assessment as not only computerized-adaptive but having the ability to be aligned with state standards and testing guidelines and as a test that measures internal growth over the course of 1 year, meeting the NCLB criteria for growth (a year's worth of growth in a year's time).

Reports and Measures

RIT. Assessments developed by NWEA use a scale called RIT to measure student achievement and growth. RIT stands for Rasch unIT, a measurement scale developed to simplify the interpretation of test scores. The RIT score relates directly to the curriculum scale in each subject area. It is an equal-interval scale, like feet and inches, so scores can be added together to calculate accurate class or school averages. RIT scores range from about 100 to 300, depending upon the scale and test season. They make it possible to follow a student's educational growth from year to year (NWEA, 2011b).

Because the assessments are adaptive and the test items displayed are based on student performance rather than age or grade, a score is independent of grade-level parameters. For example, a third grader who received a score of 210 and a fourth grader who received a score of 210 are learning at approximately the same instructional level. The fact that the RIT scale is grade-level independent allows growth to be measured (NWEA, 2011b). Information in the instructional planning tools within the MAP

assessment system allow the teacher to access discrete content skills that are most appropriate for students with that score, regardless of their grade level (NWEA, 2011b).

Data from the MAP assessment are retrieved by teachers and other educational professionals from the NWEA partner database and the NWEA homepage. These data are available in many different report forms. The primary reports used by teachers are those that detail the individual and class information for their students. These are the reports referenced in the Teacher Beliefs survey instrument used in this study.

Normative data chart. In 2011, NWEA conducted a large-scale study of test records of K-11 students in the U.S. Of the 5.1 million students in the testing pool, at least 20,000 at each grade level were randomly selected for analysis to determine growth and status norms that most closely approximate the U.S. school-age population (NWEA, 2011a). This information is referenced in Appendix A.

These data provide information on mean RIT scores (i.e., scores considered to be at the 50th percentile compared to the normative referenced sample) for both math and reading. The chart shows "Beginning-of-Year Mean," "Middle-of-Year Mean," and "End of-Year Mean" values for kindergarten through 11th grades. This number is also found on grade-level and class reports for easy comparison to an individual class or entire grade level. For the purpose of this study, the end-of-year mean for math and reading is used in order to calculate the percentage of students in a grade level that scored at or above the 50th percentile. This is used as the operational definition of proficiency.

Grade-level report. The grade-level report is an aggregated list of the MAP assessment information for all the students in a specific grade level at a specific school. This report gives information on the number of students scoring at or above the mean, the total number of students assessed, and their overall RIT score and percentile.

Additional information on the report that is not used in this study includes the length of the testing session, score ranges on individual strands of concepts within the assessment, and Lexile for the reading assessment. The report gives a list of students in alphabetical order (though sorting the data by RIT is also an option). Teacher name is not shown on this form. For the purpose of this study, this form is used to gather MAP data related to the proficiency construct since that information was not needed at the individual teacher level. The percentage of students at or above the mean was compared to the survey responses for participants affiliated with that school and grade level.

Achievement status and growth report. The achievement status and growth report contains information for an individual class of students. Student names are affiliated with RIT scores from the time frame indicated. The growth prediction index for each student is listed, indicating the number of RIT points a student is expected to grow from one benchmark to another. The actual student growth index is reported, along with a nominal yes/no, indicating whether or not the student met/exceeded or did not meet expected growth. The numerical growth index indicates how far +/- the student's RIT was from the projected score. A score of "0" would indicate that the student met exactly the number of growth point expected (NWEA, 2011b). This report can be looked at in multiple forms (i.e., fall to spring or fall to fall). For the purposes of this study, the report option selected is from fall 2014 to spring 2015. This report cannot be generated by grade level; so for this study, the researcher engaged in a precise methodology to aggregate the data.

Instructional planning. Though the above-mentioned reports are essential to building-level planning and monitoring of student progress, the reports most commonly used by teachers on a frequent basis are the ones that provide information about their

specific students (personal communication). The following reports are available at the teacher level only and are the primary resource for teachers in instructional planning and decision making.

Class report. The class report contains all of the same information as the grade-level report explained previously. The difference is that this is the report that can be accessed by the teacher about his/her specific class of students. Teachers can only access reports on students for which they are considered "teacher of record," meaning they cannot access the grade-level report option (NWEA, 2011b).

Class breakdown report. The class breakdown report is also accessible by the teacher or record for specific students and contains information only about those students. This report is mainly used to identify groups of learners by similar RIT band scores on the strands within each content subject of a MAP assessment (i.e., numbers and operations or nonfiction text). The strands correspond to Common Core State Standards for the district in this study. This report separates the bands into 9-point intervals and lists the name of each child scoring within a band in a spreadsheet-like "cell." Teachers can use this information to inform instructional practices and to identify groups of students who show similar strengths and weaknesses for targeted instructional decisions.

Descartes Continuum of Learning. The Descartes Continuum of Learning translates assessment scores into skills and concepts students may be ready to learn. It orders specific reading, language usage, mathematics, and science skills and concepts by achievement level. The skills and concepts align to the goal structures and content of a state's standards. For easy reference, the skills and concepts are grouped along the continuum according to the RIT measurement scale (NWEA, 2011b, 2012). For the purposes of this study, the Descartes Continuum of Learning was referenced specifically

by name in the Teacher Beliefs survey (see Appendix B). Teachers in the sample district have this available as an instructional planning tool that can be used to identify discrete skills that a student is ready to learn, has partially mastered, or has yet to master (NWEA, 2012).

Summary

The literature on efficacy in the practice of education indicates that there are many different components within the construct that impact students. The comparisons have been examined between student achievement and overall PE, TE, and collective efficacy. It is clear that researchers have linked factors internal to teacher beliefs and abilities to student outcome measures on assessments through a variety of models. Though researchers agree that strong associations between efficacy constructs developed from SCT and student performance exist and that there are clear graphical diagrams of these connections tied to assessment, what is missing from the literature is an analysis of specific attributes within the construct. Also missing from the literature is an analysis of the possible correlation between teacher capacity and attitude. Teacher beliefs specifically related to the use of MAP as a formative interim assessment that could inform educational decision makers is not present in the literature relevant to student performance.

Chapter 3: Methodology

Restatement of Purpose

The purpose of the study was to examine associations between teacher capacity to use and attitudes towards the formative data from MAP assessments and student performance on the assessments in terms of proficiency and growth. The study sought to answer the following research questions in order to identify more specific constructs within TE that impact educational practice.

1. What is the association between teacher capacity in using MAP formative data and student performance?
2. What is the association between teacher attitude toward using MAP formative data and student performance?

Description of Participants

The research was conducted at four elementary schools in a large, urban school district in the southeastern part of the U.S. The school district's total enrollment is 145,363 students, 108,256 of whom are in elementary or middle school and currently being assessed using the MAP assessments (District website, 2014). According to the 2014-2015 20th day report, the student sample size in this study was 2,598 students enrolled in Grades K-5 (District website, 2014). Student racial demography for the district that was sampled in this study reflected the following percentage breakdown of major groups: Asian, 5.5%; Hispanic, 19.4%; Black, 41.2%; White, 30.8% (District website, 2014; 2013-2014 Grade/Sex/Race Report). These elementary schools were chosen because their combined racial demography was similar to the district as a whole; a combined average of racial makeup is as follows: Asian, 5.2%; Hispanic, 30.4%; Black, 36.9%; White, 24.3% (District website, 2015; 2014-2015 Grade/Sex/Race Report). The

students in each school participated in the MAP formative benchmark assessments as a district mandate. The schools varied in the number of years that their students and teachers have participated in the assessments from 2-3 years (personal communication).

Table 1

School Length of Participation in MAP Assessments

School	Years of MAP Assessment
A	2
B	3
C	2
D	2

One school participated in a pilot of the MAP assessments for 1 year prior. The researcher chose not to make inclusion or exclusion of participation in the pilot assessments a delimitation of the study. Any variation in data collected for the school associated with the pilot is explained in the analysis. Surveys were sent to every classroom teacher of record at the four schools in kindergarten through fifth grades. A total of 115 teachers were asked to participate in the survey. The district has a total of 9,180 certified teachers. Student benchmark data from all 115 teachers were collected, regardless of their participation in the survey. Participation in a focus group was requested, using a maximum variation purposive sample design. This design was not randomized but focused on creating a sample of a population that could assist in answering research questions posed in mixed-methods research (Laerd Dissertation, n.d.). In this study, this technique was used to integrate qualitative perspectives from a sample of the teacher population to examine trends and themes in the survey responses that illuminated the results of the quantitative portion of the study. This method was used in

order to generate four focus groups with six participants in each, one at each school composed of one teacher from each grade level kindergarten through fifth grade, a maximum variation sample of survey participants. This heterogeneous sampling allowed for a range of perspectives to emerge which helped uncover themes and trends present in the survey data and in the sample populations (Laerd Dissertation, n.d.). It was necessary to set this criterion for focus-group participation because the data collected were aggregated by school and grade level.

Description of Instrumentation/Measurement Procedures

A survey was developed by the researcher using Google Forms (see Appendix C). The survey contained four demographic questions: (1) indicate the elementary school where you currently teach, (2) grade level currently teaching, (3) range of years teaching, and (4) number of years teaching in a school using MAP assessments. School and grade-level information was used to identify the respondents' data as part of a specific set. Years teaching and years using MAP data were asked for the purposes of examining potential trends that might impact the results of the survey responses. Three of the items were closed, multiple-option questions. The initial question asked the respondent to type in a text box the 3-digit code of their current elementary school in order to protect the anonymity of the other school sites participating in the study. Eighteen Likert-scale items were asked: 12 using a 5-point scale where 1=Strongly Disagree, 2=Disagree, 3=Neutral, 4=Agree, and 5=Strongly Agree; and six using a 5-point scale where 1=Never, 2=Rarely, 3=Sometimes, 4=Often, and 5=Frequently. For the purposes of collecting data on both research questions, the items were either designated as related to teacher attitude or teacher capacity. By design of the researcher, that information was not made available to the participants, so no section headings were used. Six items were asked related to

teacher capacity to use MAP formative data for instructional purposes, all in the agreement format. Twelve items were asked related to teacher attitude towards using MAP formative data for instructional purposes, six in agreement format and six in frequency format (see Appendix D).

The survey was developed independently by the researcher. The researcher obtained permission from Dr. Megan Tschannen-Moran to modify the TSES (Tschannen-Moran & Woolfolk Hoy, 2001). Permission to use this survey can be found in Appendix E. In order to better answer the specific research questions in this study, that scale was used as a reference and consulted for content and item structure. The researcher also consulted the TES (Gibson & Dembo, 1984) about the structure of teacher belief items. The researcher used a team of colleagues with survey development understanding and/or expertise in the education field to provide feedback. The initial iteration of the survey was sent to four colleagues. Their recommendations for changes included (1) allow the participants to choose more than one grade level; (2) switch item numbers 6 and 7 because the 5-point scale changed from level of agreement to frequency; (3) add a line in the directions that indicates that both levels of agreement and frequency were being measured and add descriptors to numbers 2, 3, and 4 on the scale; and (4) invert the choice order to be ascending for the demographic items 3 and 4. The recommended changes were made to the directions and items 6 and 7 inversion because the researcher agreed those changes were needed to clarify the survey items. The researcher chose not to make the changes recommended to invert the choice order of the demographic items or allowing for multiple grade levels to be chosen. The former was determined to be "typical" with other demographic questions that the population responds to on a frequent basis. The latter was not changed because it would cause an error in data collection that

would impact the analysis of the data which requires aggregation by grade level. The other experts agreed with the researcher that the use of the word "primarily" in the item "Which grade level do you primarily teach," prompted the respondent to choose the grade level with which they most strongly affiliate. Participants choosing multiple grade levels would result in an error in instrumentation. The researcher made these changes and sent the survey to two additional experts for review. No recommendations for changes were made by those experts in the second iteration. The survey was sent to 33 elementary teachers in a nonparticipating school within the same district to collect pilot data. These data were input into SPSS to obtain a reliability coefficient (alpha level) for the entire survey and both constructs separately.

The response rate for the pilot survey was 16/33 or 48%. The alpha level for the TSES was .90 on the short form which most closely approximates the length of the survey used in this study (Tschannen-Moran et al., 1998). However, social sciences accept that an alpha level of .70 is considered acceptable with .80 being considered good reliability (George & Mallery, 2003; Tschannen-Moran et al., 1998). For this reason, the researcher set the alpha level for the Teacher Beliefs survey developed for this study at .80. It was noted by Tschannen-Moran et al. (1998) that a factor analysis revealed that there were two distinct item types on their scale, TE and PE. For this reason, they recommend not computing an overall score because two different constructs are being measured. This supported the researcher's decision to compute statistical associations separately on teacher attitude and teacher capacity items.

Alpha levels were obtained on the Teacher Beliefs survey in its entirety to determine if the two constructs measured were related to each other (.934). Individual alpha levels were obtained on the two independent constructs—teacher capacity to use

MAP data (.953) and teacher attitude toward using MAP data (.944) (SPSS). The alpha levels obtained on all three constructs met the criterion set forth by the researcher and are considered to demonstrate good reliability. Some research calls into concern alpha levels that are extraordinarily high (such as those obtained by the survey in this study). This can indicate that there is redundancy among items on the survey. The researcher decided that because of the limited research on teacher capacity and attitude related to this topic, redundancy would seek to strengthen any associations found in the analysis of the data, because respondents should report similarly on like items.

Research Design

An embedded mixed-method research design was used to explore perceptions from a cross section of participants in a focus group to deepen the understanding of the larger quantitative study that explored associations between teacher beliefs and student performance. This design, which arose in the 1980s as a way to integrate both research methodologies, has practical applications in education where true experimental research cannot always be achieved. Using embedded mixed-methods designs strengthens the overall findings of research studies by combining both statistical results with qualitative perspectives of the participants that may be used to unearth themes and associations in the quantitative data (Creswell & Plano-Clark, 2011). Creswell (2007) advocated for this type of research when access to both forms of data are available. Many types of mixed-methods research exist (Creswell, 2007); however, this study utilized qualitative research within a larger quantitative study. Hanson, Creswell, Plano Clark, Petska, and Creswell (2005) used the notation QUAN(qual) to describe this design, indicating that the qualitative component was nested within the quantitative study (Hanson et. al, 2005). This study focused on a quantitative design comparing summed constructs on teacher

survey responses to student performance on the MAP assessment. Focus groups provided a method to explore themes and trends in a qualitative format via participant discussion of the survey response data from their school (Laerd Dissertation, n.d.).

Threats to validity. Threats to internal and external validity are noted by the researcher. The survey instrument was created by the researcher and piloted, but it is a new measure with no previous research to support its validity. In order to address this, the researcher used SPSS to obtain alpha levels on the pilot data (.934) that exceeded the threshold for "good internal consistency" (.8). The researcher mirrored other surveys that have been widely researched in the creation of the new instrument. In addition, alpha levels for the actual data set were calculated to analyze the internal consistency of the measure with a larger sample of data.

The participants self-reported their perceptions of their own capacity and attitude, which can be seen as a possible threat to validity. It is considered an assumption of this study that the participants responded accurately to all Likert items. In order to increase the likelihood for accurate self-reporting, the survey data were collected anonymously, with the exception of the school and grade level of the participant, which were known only to the researcher. This was explained to the participants in the directions of the survey and in the email that delivered the survey. The data analysis was conducted at the grade school level only. It is acknowledged that by chance some of the participants of the focus groups may have been familiar with the researcher. For this reason, a third-party facilitator was used to conduct the discussions.

External threats to validity exist in a purposive sample because it is a nonrandom sample which may or may not approximate the population from which it was extracted. Statistical analysis was used to determine how generalizable these data are to other

groups.

Operational definitions of variables. Associations were calculated between independent variables (teacher beliefs) and dependent variables (student performance). The researcher identified two constructs within each of the variable types. The independent variables are defined as

(x_1) = the mean of the scores of the Likert items related to the capacity construct.

(x_2) = the mean of the percentage values of Likert items related to the attitude construct.

The dependent variables are defined as

(y_1) = the percent of students who scored at or above the grade-level mean (50th percentile) on the MAP spring 2015 assessment.

(y_2) = the percent of students who met their MAP growth goal on the MAP spring 2015 assessment.

The dependent variables were calculated separately for both of the MAP assessment contents—math and reading. The independent variables remained the same for both math and reading.

Limitations and delimitations. Limitations of this study included that in order to protect participants, the association data were analyzed and reported in aggregate form at the school grade level. Focus group discussions allowed for the survey responses to be discussed among a cross section of members of the school. This data allowed the researcher to identify trends, themes, or actual reported reasons for survey responses which attempted to override the limitation. Limitation is also noted in the sample size and selection method of participants as a nonrandom sample of the district in which the study occurred. The researcher attempted to account for this limitation by choosing

schools with a composite similar makeup to the district population in order to approximate a normal distribution. Also, additional statistical analysis of the association between the variables was conducted using Spearman's *rho* in order to determine the level of predictive value and generalizability to the population appropriate from the sample data that were collected.

Delimitations of this study included the researcher's decision to ask survey questions only about teacher capacity and attitude, acknowledging that there are other factors that can influence student performance. Purposive sample selection was also noted as a delimitation of the study. The researcher obtained a sample that was demographically similar in student makeup to the population of the district at large in order to approximate a normally distributed sample.

Description of Procedures

Surveys were sent to the 115 teachers at the four elementary schools that were chosen as the sample. The survey was created using an electronic survey form. It was distributed to the teachers through their school email from the researcher who is also an employee of the school district. The accompanying narrative identified the survey as part of the doctoral dissertation process of the sender which was voluntary but approved by both the school district and the building principal. The survey was opened for a period of 4 weeks, with reminders sent after weeks 2 and 3. At that time, the survey response rate was 43% which approximated the pilot survey response rate and exceeds the acceptable response rate for social sciences research. Survey responses were collected into a spreadsheet that was anonymous.

Focus group members were selected using a purposive sample design, where the researcher attempted to minimize bias by putting the names of the teachers for each

school and grade level on slips of paper and having a third party select a name. That process was repeated for each school and grade level until six names for each school were selected. Those participants were contacted in the same manner as the survey in an email that was blind-copied. This was done at the design of the researcher to minimize participant knowledge of focus group members. The researcher gave the requested participants 5 business days to respond and then sent a follow-up email. If no response was returned within 2 additional business days, another name from the same school and grade level was selected using the original selection procedure explained. This procedure was repeated as necessary until all members of the grade level at each school were asked in an attempt to obtain a focus group composed as designed by the researcher to include one teacher per grade level.

Focus groups were conducted at each of the four elementary school sites 1-4 weeks after the end of the survey. Each focus group lasted approximately 45 minutes at the direction of the school district as part of their internal educational research policies and guidelines. A conference room at each school was used to provide a comfortable environment for the teachers. This space allowed participants to sit comfortably and still create an intimate environment for discussion among the participants. A third party facilitated the discussion among the participants using questions and protocol generated by the researcher (see Appendices F-H, in order of occurrence). The researcher chose to use a third party who had less intimate knowledge of the subject matter and no personal relationships with any participants. This was done to limit bias or judgment of the answers of the participants by the facilitator who may have been associated with the researcher who is employed by the same district, has intimate knowledge of the subject matter, and may have been known to some of the participants. Audio recordings of the

session were taken on two different devices to ensure that all information was captured for review by the researcher. Protocol for questions posed by the third-party facilitator on behalf of the researcher asked the participants to comment on how respondents to the survey answered on each variable of teacher attitude and teacher capacity related to the use of MAP formative data.

Data were collected from the NWEA local partner database. MAP data related to proficiency were obtained by running a grade-level report for each grade level, kindergarten through fifth, at each of the four sample schools in both mathematics and reading. That report was used to obtain the percentage of students at that grade level whose RIT was at or above the mean compared to a national normative sample. This mean equates to the student scoring at or above the 50th percentile. According to NWEA (2014), cut score percentiles for Grades 2-5 indicate that scoring in the 50th percentile would correlate with receiving at least a Level 3 on the North Carolina EOG summative assessment in both math and reading, which is the minimum level considered proficient by the state guidelines (NWEA, 2014). A sample of this report with information referenced above can be seen in Appendix I.

MAP data related to growth cannot be extracted from the NWEA database by grade level. The researcher obtained achievement status and growth reports for each teacher associated with a particular grade level in each of the four schools. Student growth was reported in two ways on the report—as yes/no nominal data on the report by student separately in mathematics and reading and by growth index, where 0 indicates that the RIT score from fall to spring of that academic school year has increased exactly the same as the student's predicted growth value as calculated automatically by NWEA within the MAP program. Negative and positive growth indices indicate how far +/- the

student's RIT score is from the predicted growth value (i.e., if a student scored 202 on the fall administration of the MAP benchmark and had a predicted growth value of 11 and then scored a 213 on the spring administration, a growth index of 0 would be recorded on the report; nominally a "yes" would be reported). A sample of this report with information referenced above can be seen in Appendix J.

For the purposes of this study, the researcher used the aggregated growth data for each teacher which is calculated by NWEA and reported on the same report. This value is reported as the percentage of students who met their growth goals as calculated by MAP. This value was calculated separately by class and subject (math and reading). For departmentalized grade levels (i.e., teachers who teach two groups of students the same subject), one composite percentage was calculated by the researcher by adding the percentage reported for each class and dividing by two to obtain the average percentage of that teacher's students who met their growth goals. If the teacher taught both subjects, then the percentage reported for each subject was used in the grade-level aggregate. Once each teacher's growth percentage was identified, the researcher calculated an average for each grade level in each school. This value indicated the percentage of students who met their growth goal in each grade level, delineated by subject. A sample of this report with information referenced above can be seen in Appendix J.

Data Analysis and Display Procedure

A mixed-methods approach was used resulting in the collection of both qualitative and quantitative data. Three types of quantitative data were collected—interval, ordinal, and nominal. Employing an embedded design allowed the researcher to utilize the data independently as opposed to converging the two types. Prior to the analysis of the data, the researcher used SPSS to run a Cronbach's alpha on the 18 Likert

items from the survey, just as had been done in the pilot. Data from each of the four schools was run separately and then together. This method was used to determine differences in reliability of the survey data among schools. This information accompanied data analysis of trends and themes that emerged from the focus group discussions. These data were used to determine if the data from the actual sample met the reliability coefficient of .80 set by the researcher as the criterion for good reliability. The data collection and analysis were conducted in phases—quantitative data from the survey, followed by qualitative data from the focus groups, then quantitative data from MAP assessments. Comparisons of the qualitative data from the survey and the quantitative data from the MAP assessment were then analyzed for strength of associations using Pearson's *r* and Spearman's *rho*.

Phase 1 involved the collection of survey data that the researcher quantified by combining the six Likert items that measure teacher capacity to obtain scores for one independent variable (x_1= teacher capacity) for each participant. For teacher capacity, the score range is 6 to 30, where each individual item has a maximum score of 5 and a minimum score of 1. Capacity items were scored on a maximum/minimum scale to obtain a self-report of overall capacity in using MAP data. Using this method allowed for compensation of strength and weakness outliers that may not have been reflective of overall capacity. The score of each Likert item was added together to produce an overall level of capacity. Using the Rasch Model, the researcher set the criterion at 25 or more for "highly capable" and 8 or less for "highly incapable."

Next, the researcher quantified the 12 Likert survey items related to teacher attitude (x_2). The researcher chose not to use the same method as the teacher capacity items because the score method computes an overall capacity, allowing for strengths and

weaknesses among the individual items. The researcher believed that getting an overall attitude score would be a less accurate reflection of the construct than other methods. For this reason, the research counted the number of survey items where the participant responded "favorably" (i.e., strongly agree, agree; frequently, often). That number was divided by the total number of items (12) to obtain a percentage of "favorability." Using the Rasch Model the researcher set the criterion at 10 of 12 (83%) or more for the "highly favorable" description and 3 of 12 (25%) or less for the "highly unfavorable" description.

The "levels of capacity" and "levels of favorability" terms were used as descriptors for the narrative analysis of the data. The criterion percentages for both levels approximated the polytomous Rasch Model which according to Andrich (2005), can be used to approximate thresholds for specific scores in Likert scales when the integer scores remain in their natural order (i.e., 1 is strongly disagree, 2 is disagree, etc.) to represent levels of a trait (Andrich, 1978, 2005).

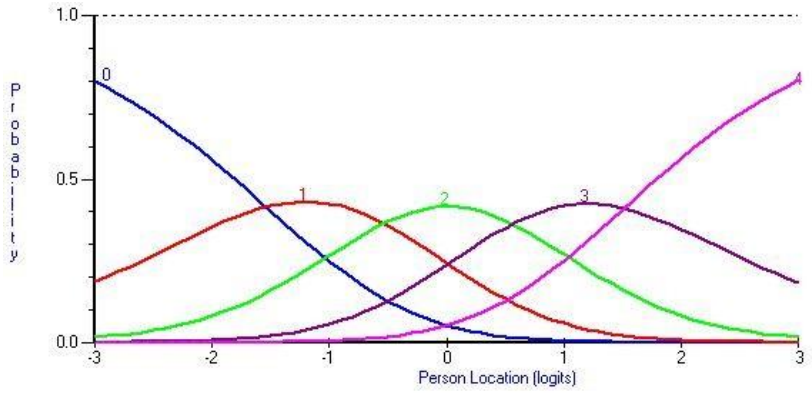


Figure 5. *Rasch Category Probability Curves for an Item with Five Ordered Categories.*

Based on the model above, the probability of a participant selecting either of the two highest values in the scale or the two lowest values in the scale is +/- 1 standard deviation from the mean of the playtokeurtic model (Andrich, 1978). In terms of

percentages, this closely approximates a normal distribution, where only approximately 34% of the population would be represented outside +/- 1 deviation from the mean. Dividing that value in half, a value of 17% should be closely approximated to be represented at either end of the data set (i.e., choosing the highest two or the lowest two values with greater probability). For the purposes of this study, the researcher set criterion for capability and favorability at the 83rd percentile (100-17= 83) on the positive end of the continuum and approximately the 25th percentile on the low end of the continuum. It is of note that the lower threshold is slightly higher than assumed by a normal distribution because any lower than that would assume that some items were omitted or that the same selection of 1 was chosen for all 18 items.

For the purposes of use with the focus group participants, the percentage of response rates for each item and category were calculated into percentages and frequencies by school and grade level. This information was used to structure the focus group protocol questions that were presented to the participants for their discussion.

In phase 2, a cross section of survey participants was assembled at each school to discuss the results of the survey. The attempt of this part of the study was to use the qualitative data obtained from the conversations to draw conclusions about why participants responded in a particular manner. This level of data served to deepen the understanding of associations that were found in the quantitative analysis. The focus group data were limited to the results of the survey and did not include information about MAP data at the design of the researcher. The transcripts from the focus groups were read by the researcher for initial overall impressions and then reread to identify emerging themes which were coded for comparison with transcripts from all four focus groups (Creswell, 2013). Creswell (2013) included Rossman and Rallis's (2012) definition of

coding as "the process of organizing data by bracketing chunks and writing a word representing a category in the margins" (pp. 197-198). In this study, the researcher used Tesch's (1990) Eight Steps in the Coding Process. This allowed the researcher to identify emerging themes directly from the focus group participants, compare transcripts for commonalities, and choose a descriptor for each of the themes that was reported for frequency of occurrence by school in table form (Creswell, 2013; Tesch, 1990). In order to connect themes to the quantitative data from the survey and MAP data, the researcher looked for connections among themes in areas of teacher attitude and teacher capacity (efficacy constructs in this study) and student proficiency and student growth (student performance constructs in this study). Since this was an embedded mixed-methods QUAN(qual) research design, the final step in this process was to analyze the themes and constructs of the study to offer insight into the quantitative results of the study (Creswell, 2013).

In phase 3, the quantitative data from the survey and the quantitative data from the MAP assessments were aggregated to the school and grade level. Since the survey instrument consisted of more than four Likert items that were summed to obtain a composite measure of a specific construct (i.e., capacity or attitude), Boone and Boone (2012) suggested that it is most appropriate to analyze this data at the interval measurement scale. It is suggested that the appropriate statistical test for measures of association for Likert scale data is Pearson's *r* (Boone & Boone, 2010). Associations were calculated among both constructs of teacher beliefs and both constructs of student performance data. Strength of association was calculated between variables using Pearson's product-moment coefficient, *r*. Separate calculations were run between x_1 and y_1; x_1 and y_2; x_2 and y_1; and x_2 and y_2 (Social Science Statistics, n.d.).

Research indicates that *r* values ≥ +/- 0.5 are considered to demonstrate large associations between variables (Laerd Statistics, n.d.). According to UNESCO (n.d.), when data are considered to be nonlinear or have a nonnormal distribution, a Spearman Correlation (*rho)* should be calculated in addition to the Pearson Correlation *r* because relationships between the variables may be nonlinear but monotonic. Spearman's *rho (R)* is intended to deal with continuous, nonnormal data sets by using rank order to account for the lack of randomness in the sample (Laerd Statistics, n.d.). Laerd Statistics (n.d.) stated that this follow-up should be conducted with data that may be nonlinear or have nonnormal distribution to see the comparison between *r* (Pearson) and R (Spearman). If R>*r*, then the data set is monotonic, nonlinear data which will show an association between the variables but not a linear, predictive relationship (Laerd Statistics, n.d.).

Data are displayed using tables, with disaggregated variable associations individually displayed. Data are displayed by school, grade level, and as an aggregate of the four schools showing the associations between variables as defined. Each construct—teacher attitude and teacher capacity—was considered separately.

Summary

The QUAN(qual) study examined the association between two specific constructs of TE—teacher capacity and teacher attitude—and student performance. Student performance was also decomposed into proficiency and growth, since both are components of student achievement data reported by the MAP assessment (NWEA, 2012). Participants responded to a survey containing Likert items related to efficaciousness. Those data were compared to grade-level proficiency and growth percentage student data on the spring 2015 MAP assessments in both reading and math to determine the strength of the association between the teacher beliefs and student

performance. Focus group participants were asked to discuss survey results for particular Likert items and both teacher capacity and attitude constructs to deepen the explanation for certain responses or trends in their school. These data were analyzed separately, using thematic coding from the strength of association data (Pearson's *r* and Spearman's *rho*) as a qualitative component that will strengthen the underlying reasons for survey response data. This research design allowed the researcher to examine both the level of association between teacher beliefs and student performance related to MAP and potential explanations about how and why those beliefs were formed.

Chapter 4: Results

Restatement of Purpose

This study examined the relationship between constructs of TE and student performance data. The purpose was to determine whether there was a relationship between teacher attitudes towards the data from the MAP interim assessment and its instructional usage and student proficiency and growth. The study compared a self-report of teacher beliefs about their capability to use MAP data and their attitude towards the data and how well their students performed on measures of proficiency and growth during 1 academic year. The goal of the study was to inform practices involving efficacy and student learning outcomes as well as to expand the theoretical construct of TE to include attitude as a separate measureable component.

Descriptive Data

Participants. As described in detail in Chapter 3, the participants of this study were 115 elementary teachers (K-5) from a large, urban school district in the southeastern U.S. Each of the participant's students participate in the MAP assessments as a mandate of the population district. The participants teach across four elementary schools within the district whose combined demographics mirror those of the district as a whole. All schools within the study had been using the MAP assessments in math and reading for either 2 or 3 years. Table 2 shows the number of teachers surveyed; number of students taking MAP assessments; and the years the school has participated in MAP testing, disaggregated by school. All schools are classified as elementary, comprised of grades kindergarten through fifth.

Table 2

School Descriptive Information

School	Number of Teachers Surveyed	Number of Students	Years of MAP Assessment
A	36	828	2
B	21	451	3
C	28	629	2
D	30	690	2
Total	115	2,598	

Survey Data

The Teacher Beliefs survey that was created for this study was used to survey the 115 teachers described above. Pilot data on the survey instrument were collected and validated with a Cronbach's alpha of .934 on how closely the constructs of attitude and capacity items were related to each other. Independently, the alpha levels were .953 for capacity items and .944 for attitude items. The data from the actual survey responses were also validated using the same measure. The alpha level for the actual survey response data was .921 for overall relationship among items. Considered separately, the alpha level for capacity was .89 and for attitude was .94. These rates were statistically similar to those found in the survey pilot. These Cronbach levels exceeded the threshold of .8 set by the researcher for validity.

Demographics. Table 3 reports the number of survey respondents by school and indicates the corresponding response rate. These response rates were similar to the pilot survey response rate of 48%. The combined response rate was 43%. It is noteworthy that School C had a significantly lower response rate than the other three schools.

Schools A, B, and D had similar response rates to both the survey and the focus group. The researcher was aware of no differences in methodology that led to this disparity with School C. The number of participants willing to attend a focus group was significantly lower than the desired number of six per group described in the methodology. Again, no reason that is known to the researcher led to the low number of focus group respondents.

Table 3

Survey Response Rates

School	Number of Survey Respondents	Survey Response Rates	Number of Focus Group Responses
A	15	42%	2
B	14	67%	2
C	4	14%	0
D	17	55%	2
Total	50		6

The Teacher Beliefs survey included demographic items that are summarized as aggregates of the four schools in the tables below. Frequency of survey respondents by grade level is indicated in Table 4. The researcher found it notable that the total number of respondents in Grades K-3 was 18 and in Grades 3-5 was 32. It should be considered that at the time of the study, the state of the population district had summative state assessment data in Grades 3-5 that was tied to teacher effectiveness data. NWEA has linking data available between MAP and the summative assessment measure in the state where the study occurred.

Table 4

Survey Respondents by Grade Aggregate

Grade Level	Number of Survey Respondents
K	4
1	7
2	7
3	11
4	12
5	9

Frequency of survey responses by level of experience is shown in Table 5. The number of responses does not vary among differing amounts of teaching experience except teachers who were in their first year of teaching at the time of the study. Statistically, it is likely there were not as many first-year teachers as other demographic experience levels measured on the survey because the population of first-year teachers in the sample schools is less than the population of other categories.

Table 5

Survey Respondents by Experience Aggregate

Years of Teaching	Number of Survey Respondents
15+	12
10-14	9
5-9	12
1-4	15
<1	2

The population district has been using MAP in some schools for up to 3 years. The survey asked participants to indicate how many years they had worked in a school

that used MAP assessments. Respondents who chose more than 3 years either had to work in a school that used MAP assessments in another school district or they were erroneous in their information of how long their school had been using the assessment data. This information was of interest to the researcher, not to answer the research questions central to this study but to inform practices, should that demography indicate statistical differences in reporting.

Table 6

Survey Respondents by MAP Experience

Years Using MAP	Number of Survey Respondents
>3	17*
3	15
2	10
1	7

Note. *11 of the respondents had experience outside of the population district.

Efficacy data. The survey instrument was divided into two parts—six items that measured capacity and 12 items that measured attitude. Teachers self-reported on a 5-point Likert scale for each item. The tables below indicate the average score by item for each school in the research study. Table 7 reports the average aggregate scores by school on the survey items that measured teacher beliefs about their capacity to access, understand, and use MAP data for instructional practices.

Table 7

Survey: Capacity Item Aggregate Analysis (Scale 1-5)

School	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6
A	4.2	4.1	4.3	4.1	3.8	3.9
B	4.1	3.9	4.4	3.9	3.6	4
C	4.5	3.3	4.3	3.3	3.8	3.8
D	3.9	3	4.1	3.9	3.9	3.6
Average	4.2	3.6	4.3	3.8	3.8	3.8

Table 8 reports the average aggregate scores by school on the survey items that measured teacher beliefs about their attitudes towards the MAP assessment, data, validity, and measures of learning for their students. The researcher indicated a note that questions 8, 12, 15, and 16 had significantly lower average aggregate scores. Those items were the items most closely associated with how teachers feel about the MAP assessment data as a measure of their students' learning and use of the MAP data for instructional decision making (Table 8).

Table 8

Survey: Attitude Item Aggregate Analysis (Scale 1-5)

School	Item 7	Item 8*	Item 9	Item 10	Item 11	Item 12*	Item 13*	Item 14	Item 15*	Item 16*	Item 17	Item 18
A	4.2	2.7	4.5	3.9	4.3	2.8	3	3.3	2.9	3.1	3.7	3.6
B	4.2	3.6	4.2	4.3	4.2	3.3	3.4	3.7	3.2	3.6	4	3.8
C	3.5	2.5	4.3	3.5	3.5	2.8	3.8	3.3	2.8	3	3.3	3
D	3.8	3	4.2	4.1	3.6	2.7	3.1	3.4	2.9	3.2	3.5	3.4
Average	3.9	3	4.3	4	3.9	2.9	3.1	3.4	3	3.2	3.6	3.5

Note. *Statistically lower than other items.

As described in Chapter 3, scores for capacity and attitude were calculated based on the average ratings by respondents. The capacity score was reported as "out of 30,"

and the attitude scores were a "percent of 100." School data were averaged to gain an overall capacity and attitude score for each participating school. These scores are depicted in Table 9. School C, though it has a comparable capacity score, differs from the other schools in its attitude score. It is noted that there was a low response rate at School C.

Table 9

Capacity and Attitude by School Aggregate

School	Capacity Score (of 30)	Attitude Score (% of 100)
A	24	42.4%
B	24	62.5%
C*	23	29.2%
D	23	53.6%
Average	23.5	46.9%

Note. *Response rate < 15%.

In order to determine if differences in capacity and attitude scores existed among grade levels, Table 10 indicates the average of the four schools' capacity and attitude scores by grade. Noted was that kindergarten and fourth grade had only two schools with respondents in those grade levels. First grade shows a slightly lower average attitude percentage, but no statistically significant differences were noted among grade levels in either capacity or attitude. School B had the highest capacity and attitude ratings. A teacher at School B had this to say about a possible reason, "we use MAP a lot also to drive our instruction; even though it is above 50 percent [attitude percentage], I would still expect our school to be higher because they push MAP so much" (Focus Group School B, personal communication, 2015).

Table 10

Capacity and Attitude by Grade Level Aggregate

Grade Level	Capacity Score (of 30)	Attitude Score (%of 100)
K*	25	68.6%
1	22	35.8%
2	23.25	46.4%
3	24.5	67.8%
4*	23.7	57.1%
5	23.25	41.2%

Note. *Only 2 schools reporting.

The researcher outlined descriptive terms for capability and favorability in Chapter 3 based on the Rasch Model for interpreting Likert scale models. Table 11 shows the percent of the respondent scores that fell within the ranges of the defined terms. This information is reported by school average. Disparity existed among percentages of respondents who felt they were highly capable of using MAP data for instructional decision-making practice. Grade 3 showed a high percentage on the attitude construct. A teacher at School B surmised that "maybe it has a lot to do with the Reading to Achieve [law governing 3rd grade Reading pass rates for EOG assessments]" (Focus Group School B, personal communication, 2015). Data from School C is disparate from the other data reflected in the table; but it should be noted again that this school had only four respondents, a response rate of less than 15%. These data are not considered to be a reflection of the school teacher perceptions. School D showed a significantly lower perception of respondents who believed they were highly capable in using MAP assessment data in practice. Each focus group was asked to look at overall attitude percentages, and a teacher at School D made the observation that "[most] of the grades

kind of sit right in the middle . . . which kind of shows me like some uneasiness about their attitude about it" (Focus Group D, personal communication, 2015).

Table 11

Capability and Favorability

School	Highly Capable	Highly Incapable	Highly Favorable	Highly Unfavorable
A	40%	0%	0%	0%
B	50%	16.7%	33.3%	16.7%
C	25%	0%	0%	75%
D	16.7%	0%	16.7%	16.7%
Average	32.9%	4.2%	12.5%	27.1%

Note. Reported as percent of grade levels.

When asked about this data at a focus group at School A, both participants commented about a possible reason for the moderate capability score and the lack of favorability, stating, "I think that we have not learned how to use the data probably like we should" (Participant 1); and "I think it's a lot of pressure and stress . . . those kids better show growth" (Participant 2) (Focus Group School A, personal communication, 2015).

School D showed a significantly lower perception of respondents who believed they were highly capable in using MAP assessment data in practice. This was explored in the focus group for School D, where one teacher said,

> Obviously people are at that point where they know how to access what they need. Perhaps they just don't have the time, or they are not finding it valuable. Personally, I think it's a time factor. I know I would love to use that information. I think it would be very helpful to guide instruction if I had a reasonable amount

of time to utilize what is being provided for me, because I think it gives very detailed information. (Focus Group School D, personal communication, 2015)

Focus Group Data

Focus groups were formed using purposive design sampling explained in detail in Chapter 3. The design of the study was to have six participants per school, one from each grade level. After exhausting the sample, only two participants at Schools A, B, and D agreed. This was cited as a limitation of that portion of the study. A focus group was not conducted at School C because no one agreed to participate. This was a similar response to the survey response rate for School C, which was four respondents. Since this was an embedded mixed-methods QUAN(qual) study, the focus group data were used as explanatory data about the survey responses at each school. Coded themes are indicated by frequency in Table 12.

Table 12

Themes by Frequency

School	Lack of Time	Technology Concerns	Performance Pressure	Correlation	Student Efficacy	Lack of Trust/ Value	DDDM Training
A	3	2	4	6	7	2	4
B	0	0	3	5	0	2	3
D	5	5	4	8	7	7	8
Total	8	7	11	19	14	11	15

Note. Reported as number of separate times mentioned by participants.

The seven themes presented in the table above represent those with the highest frequencies. It should be noted that correlation between MAP and other assessments was the most frequently mentioned theme, followed by lack of training in DDDM with their

specific student data on how to translate into action steps and then student efficacy or student attitude towards the assessment. Since student attitude was outside the scope of this study, it is noted that performance pressure and lack of trust/value in the MAP assessment were also mentioned frequently within the focus groups. It is considered important to note that School B noted many less areas of concern. Participants 1 and 2 both reported that they "used MAP a lot" and expected attitudes "to be higher" because of the schools internal practices. School D mentioned lack of time frequently, but the research noted that of the times it was mentioned, six were connected to lack of time to spend using the data to make decisions, so some crossover can be assumed between lack of time and DDDM Training.

MAP Data

Student performance data on both proficiency and growth were calculated from MAP assessments in math and reading for each grade level at each school in the study. These data were averaged to acquire a school-level aggregate for proficiency and growth. Proficiency is defined as the percent of the students scoring at or above the normative grade-level mean set by NWEA, denoted by the 50th percentile. This is the same indicator for both math and reading across all grade levels. Growth met is the percentage of students whose scores increased *at least* the amount predicted by NWEA on the MAP assessments in either math or reading. This indicator is the same across all grade levels. It is important to distinguish that these measures are independent indicators of student performance (i.e., a student could be proficient and not meet growth or not be proficient and still meet growth). The data in Table 13 report the proficiency and growth average percentages for each school by subject. In previous tables, School C's data were notable because they were reflective of minimal survey respondents. This table is the school

aggregate unrelated to the numbers of respondents, so it is notable that School C has lower average data points in most areas compared to the other schools in the study. School D reflects particularly low proficiency data for both math and reading, though notes that growth regressed towards the mean of the other participating schools. The researcher noted that similarities between math and reading proficiency and math and reading growth-met percentages existed for each school.

Table 13

MAP Student Performance Aggregate Data

School	Proficiency (Math)	Proficiency (Reading)	Growth Met (Math)	Growth Met (Reading)
A	65.95%	66.13%	67.07%	61.04%
B	59.82%	65.33%	69.53%	63.99%
C	39.62%	47.33%	53.62%	51.46%
D	34.80%	34.25%	60.51%	54.91%
Average	50.05%	53.26%	62.68%	57.85%

Correlational Analysis

The research questions in this study sought to determine the strength of the associative relationship between teacher beliefs about their capacity and attitude in comparison to student performance outcomes—proficiency and growth on MAP assessments. Pearson's *r* and Spearman's *rho* (R) correlational measures were used to answer both questions independently. Pearson values of r>+/- 0.5 indicate a significant relationship between variables. Spearman values of R that exceed the Pearson values for the comparison speak to the generalizability of the findings to the larger population. The statistic was used since the sample was a nonrandom, quasi-experimental design.

Operational definitions of variables. The independent variables are defined as

(x_1) = the mean of the scores of the Likert items related to the capacity construct.

(x_2) = the mean of the percentage values of Likert items related to the attitude construct.

The dependent variables are defined as

(y_1) = the percent of students who scored at or above the grade-level mean (50th percentile) on the MAP spring 2015 assessment.

(y_2) = the percent of students who met their MAP growth goal on the MAP spring 2015 assessment.

The dependent variables were calculated separately for both of the MAP assessment contents—math (y_{1a}) (y_{2a}) and reading (y_{1b}) (y_{2b}). The independent variables remained the same for both math and reading.

Statistical Results

Research Question 1: What is the association between teacher capacity in using MAP formative data and student performance? Table 14 shows the correlational values Pearson's *r* and Spearman's *rho* (R) for the comparison of teacher beliefs about their capacity to use MAP data for instructional practices and student performance outcomes on the MAP assessment as measured by both proficiency and growth. None of the comparisons shown in the table below indicate a significant relationship (r>+/- 0.5) between capacity and proficiency or capacity and growth. Additional statistics were run using Spearman's *rho* to determine the generalizability to a larger sample, since the sample in this study was considered nonrandom. No statistical significance was confirmed using either measure.

Table 14

Correlational Analysis (Schools Aggregate): Capacity

Statistical Test	Capacity/ Proficiency Math (x_1 y_{1a})	Capacity/ Proficiency Reading (x_1 y_{1b})	Capacity/Growth Math (x_1 y_{2a})	Capacity/ Growth Reading (x_1 y_{2b})
Pearson *r*	.31	.24	.46	.37
Spearman *rho*(R)	.34	.31	.43	.26

Note. *Data for grade levels where survey response was 0 have been extracted.

Research Question 2: What is the association between teacher attitude toward using MAP formative data and student performance? Table 15 shows the results of the statistical analysis of the attitude variable compared to both proficiency and growth variable for both the MAP math and reading assessments.

Table 15

Correlational Analysis (Schools Aggregate): Attitude

Statistical Test	Attitude/ Proficiency Math (x_2 y_{1a})	Attitude/ Proficiency Reading (x_2 y_{1b})	Attitude/ Growth Math (x_2 y_{2a})	Attitude/ Growth Reading (x_2 y_{2b})
Pearson *r*	.27	.27	.55**	.65**
Spearman *rho*(R)	.23	.35	.45**	.57**

Note. *Data for grade levels where survey response was 0 have been extracted. **Data are considered to be statistically significant (Social Science Statistics, n.d.).

As in Table 14 above, Pearson's *r* and Spearman's *rho* were used to determine the strength of the relationship between variables. Pearson *r* values greater than +- 0.5 were

again considered to be indicative of significant relationships. Much the same as the capacity construct results above, there was no significant relationship noted between teacher beliefs regarding their attitudes towards MAP and its data when compared to the percentages of students who were considered proficient on the MAP assessment in math or reading using either statistical test.

Comparative data in the last two columns of Table 15 show that a significant relationships was found between teacher beliefs about their attitudes as measured by the Teacher Beliefs survey instrument, items 7-18, and the percentage of students who met their growth on both MAP assessments in math and reading. According to the Laerd Statistics (n.d.) calculator, (x_2 y_{2a}) had a value of r=.55, which indicates a moderate positive correlational relationship; (x_2 y_{2b}) had a value of r=.65 which indicates a strong positive correlational relationship. Confirmed by Spearman's *rho* tests, a moderate positive relationship exists between the variables of attitude and growth across multiple measures. Since R<r, as explained in Chapter 3, this nonrandom sample data cannot be generalized to a population outside this study. Further discussion of this occurs in Chapter 5.

Summary

Proficiency and growth student performance data from the MAP assessment were compared to teacher responses on the Teacher Beliefs survey instrument in the areas of capacity and attitude. These comparisons were analyzed to determine if a relationship existed between student performance and teacher self-reports of their capacity to use MAP data and attitude towards the data. Pearson product-moment correlation (*r*) tests were run to determine the strength of the relationship among the variables. Spearman's *rho* (R) tests were run on each set of variables to confirm the results of the Pearson test

and to lend strength to the generalizability of the results to larger populations, given the nonrandomness of the sample used in this study. The results found significant relationships between the aggregate teacher attitude (x2) and aggregate student growth data on both the math and reading MAP assessments (y2). No significant relationships were found between teacher attitude and student proficiency (y1) or teacher capacity (x1) and either student proficiency or growth performance. Comparison of Spearman's *rho* to Pearson *r* indicated that the data from this nonrandomized sample was confirmed to have a relationship but could not be generalized based on just this statistical comparison. Further discussion of these results is presented in Chapter 5.

Chapter 5: Discussion

Summary

The aim of this study was to address the deficiencies in the literature exploring which dimensions of TE have the most direct impact on student performance outcomes on the MAP assessments in order to inform educational practice. In addition, this study was conducted to validate other social cognitive theorists' findings about the distinct elements of efficacy. Findings support Bandura's idea that it is essential to know whether low efficacious behavior on the part of teachers is due to lack of confidence in their abilities (which this study calls capacity) or futility about their abilities to impact educational outcomes for students (which this study calls attitude) (Bandura, 1997; Soodak & Podell, 1996).

Conclusions

Findings. This study sought to determine the strength of a potential relationship between teacher perceptions about instructional practices related to the MAP assessment and student performance outcomes on the interim assessment. Research questions were separated into two constructs—teacher capacity and teacher attitude—for comparison to student proficiency and growth outcomes on MAP math and reading assessments from the spring 2015 administration.

Research Question 1: What is the association between teacher capacity in using MAP formative data and student performance? In a statistical comparison of the aggregate data of the four schools participating in the study, no significant relationship was discovered between teachers' perceptions of their capacity to access, understand, and utilize MAP data from instructional decision making and student performance outcomes related to proficiency or growth in either math or reading.

Though all the relationships were noted to be positive, none met the significance interval of r>+/- 0.5. These findings indicate that no relationship existed between how high the capacity level of the teacher was in utilizing the MAP data and student performance on the assessment at subsequent administrations. Since the majority of trainings, professional developments, and data-driven conversations about the use of interim assessment data are focused on building the skill sets of teachers to interpret data and use the information to increase student learning, teacher high perceptions of their capabilities in these areas were expected. If there is not a substantial relationship, though, between these skills and student proficiency and growth on this assessment, as indicated by the findings of this study, continuing to build data-wise skill sets in the absence of efficacious behaviors will not have a causal impact on student performance outcomes.

Research Question 2: What is the association between teacher attitude toward using MAP formative data and student performance? In a statistical comparison of the aggregate data of the four schools participating in the study, no significant relationship was discovered between teachers' perceptions of their attitudes toward the data from MAP, usage for instructional practices or measures of student learning, or student proficiency levels in either math or reading. These relationships were similar to the findings related to capacity and performance in that though the relationships were positive, they were weak or very weak associations.

Conversely, when statistical analysis of the relationship between teacher attitude and percentages of students who met their growth scores on MAP math and reading assessments were compared, statistical significance was found. The correlation was r=.55 (moderate, positive) for math and r=.65 (strong positive) for reading. These findings indicate that as teacher attitudes increase in favorability towards 100%, the

likelihood of their students meeting their growth goals as set by NWEA on the MAP assessments increases. This study concludes that attitude has a more substantial relationship to student growth outcomes than how capable a teacher is to utilize the data.

The findings of this study support the assertion that teacher attitude is a separate, measureable construct within efficacy that has a positive association on student growth. The findings support additional resource allocation towards research and educational practices that enhance teacher attitudes about interim data as a method of student growth achievement. A comment from a focus group member at School D supported this conclusion, stating,

> to truly individual[ize] and look where the student needs the most impact, and then to be able to plan and execute that—I think it's a huge time factor and that's why the attitude is [this way]. I think it would be higher if we had the data and then were given an ample amount of time to analyze and be able to utilize that information. (Focus Group School D, personal communication, 2015)

Connections to literature. The conclusions of this study increase the body of research about the expansive construct of TE relating to student learning. Researchers have continued to examine teacher beliefs in order to determine which components have substantial relationships to performance outcomes (Guskey, 1982; Labone, 2004; Tschannen-Moran & Woolfolk Hoy, 2001). Even as early as 1986, when research on efficacy was just expanding into educational science, Ashton and Webb (1986) referenced how outcome expectancies about student performance are relative to teacher judgments about their ability to impact learning. This study named those judgments teacher attitude and concluded that it is the specific construct within efficacy that is most directly related to student growth performance.

The findings of this study support the research of Soodak and Podell (1996), who indicated that the dimensions of TE were not always correlated; therefore, it is concluded that a teacher could believe he/she has the capacity to make educational decisions using available resources but lacks the attitudinal component that the data, resources, and his/her own instructional decisions bring about positive learning for students. Based on the results of this study, the attitude of the teacher is the dimension of TE that shows significant relationships to student performance outcomes, specifically growth. Understanding this impact informs educational practitioners about resource allocation and professional growth (Bandura, 1997; Bremer, 2008).

The significant correlational relationship found between student growth and a facet of TE related to MAP performance corroborates Bremer's (2008) earlier findings in a similar study. This study extended Bremer's research to examine whether capacity or attitude differed in strength of relationship to student performance on MAP assessments. Now that the validation of a relationship between MAP and TE has been demonstrated in association with teacher attitude and student growth performance, additional validation about specific efficacious constructs is necessary to develop the body of evidence.

Limitations

The major limitation of this study was the use of a small, nonrandom sample. Though the researcher attempted to counter the limitations with use of specific schools that represented the demography of the large district as a whole, the small sample size limits the generalizability of the results. The use of the Spearman *rho* statistical test attempted to give the researcher an indication about whether any of the Pearson correlational statistics could be considered generalizable to a larger population. In addition to this known limitation, one of the four schools in the sample had significantly

lower survey response rates than the other three schools. This same school did not have any participants willing to participate in a focus group. For this reason, the aggregate data is not necessarily reflective of School C. Focus groups at the other three sites were also considered limitations of the study because fewer than three people participated in each focus group at Schools A, B, and D. This limited the qualitative portion of the study which was used as explanatory data for survey responses from each school.

Implications

Theory. Theoretical implications of the findings of this study are supportive of TE as an elusive construct (Guskey, 1982) with differing components of PE and OE that are not necessarily correlated to one another (Bandura, 1997; Soodak & Podell, 1996). Bandura's (1996, 2001) modern works in efficacious behavior concluded that teacher perceptions about their abilities can differ from their attitudes about outcomes. The results of this teacher belief study confirmed that those elements of efficacy differ in their relationships to student proficiency and growth. Implications to the field are that TE cannot be studied as a single construct compared to student learning outcomes but must be addressed in many facets.

Models. In this study, two specific OE models were considered—the large-scale Value-Added Assessment Model of Balls et al. (2011) and the Educational Decision-Making Loop depicted by Masters (2013).

The findings of this study support Balls et al.'s (2011) model which indicates that the efficacy of the members of an organization have a direct relationship to the outcomes of that organization. Specifically, these findings highlight the impact of the attitudes of the members of a group as having the most substantial correlation to the results. Though not disaggregated into components of efficacy in this model of organizational OE, this

study magnifies the importance of considering how members of an organization feel about practices in order to maximize the desired results.

In regards to Master's (2013) decision-making loop for educational outcomes, the implications are more directly substantial. The major factors referenced in this model are available resources, understanding of the current practices, and knowledge of continuous improvement. Teacher capacity could be considered an available resource; however, efficacy is not considered in this model. The conclusions of this study suggest that no matter how capable educators are to impact student learning in a positive way, if they do not believe in the practice, student learning outcomes will not be impacted. As teachers increase their positive attitudes towards an assessment practice, student growth increases. This OE model for educational decision making would be strengthened by the addition of TE as an input factor into the loop. Figure 6 below is a model of the Education Decision-Making Loop with the revision mentioned in this implication.

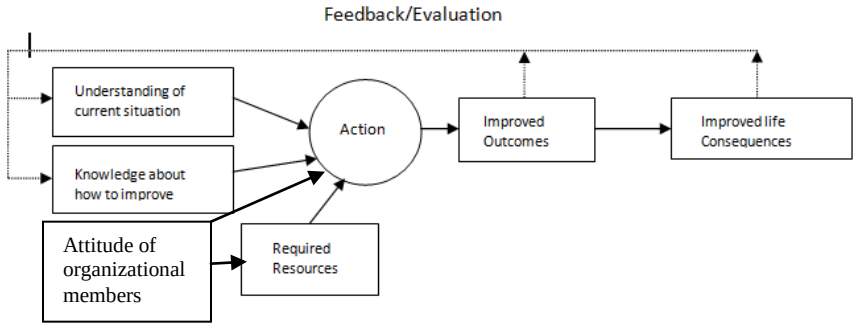


Figure 6. *Educational Decision-Making Loop (Masters, 2013) with Revisions from Implications from this Study.*

Educational practice. Implications to the field of educational practice center on how professional learning and resources are devoted to increased skill sets or capacity of teachers to utilize data from interim assessment measures to bring about increases in

student growth. This study indicates that attention to the attitudes of educational practitioners must be factored into best practices and professional development if information from these formative assessments is to be used to bring about student growth on summative measures. As educational leaders design and implement professional learning and coaching about data-driven instructional practices, collegial dialogue must occur to assess the perceptions of the teachers about the measure, its data, and its connection to student learning. Decision makers must take into account how the feelings of the practitioners using the data to make instructional decisions influence the summative results. Additional information about why certain perceptions exist must be addressed at the building level, so principals and curriculum leaders can explore how best to increase teacher buy-in to local assessment practices that aim to increase student performance. It should be the practice of district and state leaders to frequently collect attitudinal data about assessment measures that are used to evaluate teacher and student performance. The results of this study found a significant relationship between teacher beliefs and whether or not students grew the predicted amount across grade levels and content measures. These findings should be considered pertinent since teacher and district evaluations in the state where the study occurred, as well as many other states, are tied not to proficiency of students but to the growth indices of schools and teachers.

Recommendations

Population district. Results of this study indicate that a significant positive relationship exists between how teachers report their attitudes towards the data from the MAP assessments and whether or not students in K-5 are meeting their MAP growth goals on both the math and reading assessments. Though many teachers reported high capacity to use the MAP data, there was not a significant correlational relationship to

either student proficiency or growth. Based on item analyses of the Teacher Beliefs survey, this implies that when teachers believe that the data from MAP assessments is reflective of student ability, comparable to other data sources; and a good indicator of the learning in their classrooms, more of their students meet their growth goals. It is the recommendation of the researcher that the district increase professional development activities and collegial dialogue centered on the perception of MAP data and how it reflects teaching and learning in the classroom. Additional resources should be allocated to create opportunities for comparison of data across the district in order for teachers to have the ability to engage with other teachers and buildings where opinions may differ about the validity and usage of MAP data to drive instructional practices. This desire for allocation of resources was a major theme in each of the three focus groups summarized by a participant from School A as follows:

> I think we have gotten training . . . there have become experts on the staff who come back and say, "Hey, this is what you can do. This is how you can use it." Now it's just a matter of among other things taking time to sit and analyze and make groups and what does this really mean about my teaching or my students learning, you know, that kind of thing, so I think yes, we are becoming more knowledgeable in our ability to be able to attack and analyze the data, but it's just timing manpower to be able to pull the groups to do it, and do it well. (Focus Group School A, personal communication, 2015)

Based on focus group data collected from three of the schools in the sample, it is clear that teachers connect being given time to gather, analyze, and engage in facilitated conversations about their use of MAP data to improve instructional outcomes for their students and how they feel (attitude) about MAP assessments and data. These capacity

skills were cited as teachers to be some of the reasons survey respondents may have reported a less favorable attitude towards MAP data as a valid measure of student performance. Lack of time, as well as lack of training in data-driven instructional practices using their personal data, were found to be themes across all focus groups.

Since the researcher is employed in the population district, it is known that the district has provided extensive mandatory and optional training opportunities focused on the "nuts and bolts" of the MAP assessment and data system and on DDDM with MAP. It is clear from the respondents of the survey and the input of the focus group participants that this holistic training, though satisfactorily increasing the capacity of users, is not leading to buy-in from teachers about the connection between MAP and student learning outcomes. Even when teachers indicated that their personal belief was that MAP was valuable, they cited extensive numbers of mitigating reasons why that belief alone did not cause them to have a positive attitude towards the assessment. Focus group themes indicated that teachers needed to have a positive attitude about the processes surrounding how they used the MAP data and how they integrated the data into their responsibilities as instructional practitioners. They also responded that there was lack of confidence or an uneasiness about how MAP data fit into the assessment practices of their schools (i.e., all the other assessments that were given).

There was a call among participants to have guidance about the relationship between MAP data and other district data that were used to evaluate both student and teacher performance—specifically EOG assessments and Reading 3D. Focus group participants also spoke heavily on feeling that there was a lack of time to engage in professional dialogue with their teacher teams in a meaningful way. It was referenced that facilitated work with data was of interest to teachers because they need guidance on

using their personal data to drive student-centered instruction in their classrooms. Participants referenced that they knew this would lead to student growth and that should be the focus; but without the time and structure to do so, MAP seemed to be just "another thing."

Using both the statistical data from this study and central themes from the focus group data, it is the recommendation of the researcher that the study district make three specific changes to its current practices surrounding the use of MAP data. The district should utilize the NWEA Linking Study referenced in this paper and provided to the district about the relationships between the state assessment and MAP data from 2014. This information will assist schools in understanding the correlation between MAP RIT scores and EOG performance. This is powerful information so teachers can see MAP data in a context that is familiar. Since EOG data drive portions of teacher evaluation, if MAP is to be considered a valuable means of assessing student learning, teachers must understand its comparison to summative assessments. Having access to the linking information between MAP as an interim assessment and the EOG as a summative assessment will allow school-based teams to utilize a balanced data system (Perie et al., 2007) to make instructional decisions.

In addition, the researcher recommends that the district take a stronger stance on how it chooses to utilize MAP data. Teachers need a more clear answer as to whether they should focus on MAP as a proficiency or growth measure. Growth is the component that was concluded from this study to be most correlated with TE. Since prior research has shown relationships between TE and student performance (Ashton & Webb, 1986; Guskey & Passaro, 1994; Labone, 2004; Soodak & Podell, 1996; Tschannen-Moran et al., 1998), the researcher concluded that the study district should continue to focus the

use of MAP data as a means of measuring student growth relative to normative grade-level expectations.

Final recommendations to the district are considered by the researcher to be the most crucial to increasing teacher attitudes toward the MAP assessment and its data as a means of impacting student growth. Since focus group participants connected teacher attitude to how well they were actually using the skills they possessed regarding MAP information, the district should use the discretionary professional development days in the upcoming school year to assist school leaders in providing time and resources for teachers to engage with MAP data using protocols that will focus teachers on the next steps based on the student performance on the fall, winter, and spring benchmarks. Since the researcher is a member of such a team within the district, some schools are much more advanced with this type of data discussion. The district is encouraged to identify which schools have already developed a successful process for the gathering, analysis, and discussion of MAP data and its usage. Once those schools are identified, district and learning community leaders should create cross-school professional learning teams that engage in collegial inquiry. The model schools should serve in a supportive capacity to the staff of other schools as they seek to increase their data-based culture.

Within schools, principals are encouraged to create a coaching structure whereby members of the administrative and school leadership team coach other members of the staff in the analysis of MAP data to bring about instructional changes in the classroom that impact student learning.

As teachers increase their confidence in the data from MAP as a reliable source of information that helps guide them as practitioners, their attitudes towards the measure itself will increase. Based on the results of this study, that would have a statistical impact

on the number of students who meet or exceed their growth goals on MAP. Since value-added systems are being used to evaluate teacher ratings, even though MAP is not the measure utilized, it is critical that teachers are focused on practices that impact student growth.

Future research. Based on the results of this study, replication across different settings and with larger samples is needed to support the statistical findings. This study could be conducted as a quantitative, experimental design with a nationwide population and randomly assigned sample. Findings from such a study would be more generalizable and serve to confirm or deny the significant positive relationship discovered in this study between teacher attitude and student growth performance.

In addition to replication, the Teacher Beliefs survey was created and piloted for the purposes of this study. Additional use and statistical analysis of this instrument could provide further evidence of its validity and reliability. Future researchers' use of this survey with adaptations to specific assessments pertinent to their studies should indicate the same item analysis results since the constructs of capacity and attitude should be interchangeable among assessment titles. Further usage of the Teacher Beliefs survey could allow enough validity and reliability data to be generated so that this assessment could be seen as an extension of previous TE assessments (i.e., TSES) that include the new construct of teacher attitude that was defined in this study.

Final Remarks

Research continues to be necessary to define and incorporate "attitude" as a component of efficacy as defined by SCT. Increasing research around this specific construct is crucial to informing practices that involve TE as a part of their models of continuous improvement (i.e., value-added models, Educational Decision-Making Loop).

It is necessary to determine how best to allocate resources towards increasing efficacious practices that have been shown to correlate to student performance outcomes. Efficacy continues to be a complex construct that requires research that informs best practices in education.

References

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika,* 43, 357-74.

Andrich, D. (2005). Rasch models for ordered response categories. In B. Everitt & D. Howell (eds.) Encyclopedia of Statistics in Behavioral Science. New York: John Wiley & Sons. Volume 4, pp. 1698-1707.

Armor, D., Conroy-Oseguera, P., Cox, M., King, N., McDonnell, L., Pascal, A., . . . & Zellman, G. (1976). Analysis of the school preferred reading programs in selected Los Angeles minority schools, REPORT NO. R-2007- LAUSD. Santa Monica, CA: Rand Corporation (ERIC Document Reproduction Service No. 130 243).

Ashton, P. T. (1983). *A study of teachers' sense of efficacy. Final report, executive summary.* Retrieved from http://ezproxy.gardner-webb.edu/login?url=http://search.ebscohost.com.ezproxy.gardner-webb.edu/login.aspx?direct=true&db=eric&AN=ED231833&site=ehost-live

Ashton, P. (1984). Teacher efficacy: A motivational paradigm for effective teacher education. *Journal of Teacher Education, 35,* 287-232.

Ashton, P. T., Olejnik, S., Crocker, L., & McAuliffe, M. (1982). *Measurement problems in the study of teachers' sense of efficacy.* Paper presented at the annual meeting of the American Educational Research Association, New York.

Ashton, P. T., & Webb, R. B. (1986). *Making a difference: Teachers' sense of efficacy and student achievement.* New York: Longman, Inc.

Balls, J. D., Eury, A. D., & King, J. C. (2011). *Rethink, rebuild, rebound: A framework for shared responsibility and accountability in education.* (2nd ed.). Boston, MA: Pearson Learning Solutions.

Bandura, A. (1977). *Social learning theory.* Englewood Cliffs, NJ: Prentice Hall.

Bandura, A. (1986). *Social foundations of thought and action: A social-cognitive theory.* Englewood Cliffs, NJ: Prentice Hall.

Bandura, A. (1997). *Self-efficacy: The exercise of control.* New York: W. H. Freeman.

Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology, 52,* 1-26.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Education: Principles, Policy and Practice, 5*(1), 7-74.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*(1), 5-31. Retrieved from http://ezproxy.gardner-webb.edu/login?url=http://search.ebscohost.com.ezproxy.gardner-webb.edu/login.aspx?direct=true&db=eric&AN=EJ829749&site=ehost-live;http://dx.doi.org.ezproxy.gardner-webb.edu/10.1007/s11092-008-9068-5

Boone, H. N., & Boone, D. A. (2012, April). Analyzing Likert data. *Journal of Extension, 50*(2).

Boudett, K. P., City, E. A., & Murnane, R. J. (2006). The "data wise" improvement process. *Principal Leadership, 7*(2), 53-56. Retrieved from http://ezproxy.gardner-webb.edu/login?url=http://search.ebscohost.com.ezproxy.gardner-webb.edu/login.aspx?direct=true&db=eric&AN=EJ767107&site=ehost-live;http://www.principals.org/KnowledgeCenter/Publications.aspx

Bracey, G. W. (2007). The proficiency illusion. *Phi Delta Kappan, 89*(4), 316. Retrieved from http://ezproxy.gardner-webb.edu/login?url=http://search.ebscohost.com.ezproxy.gardner-webb.edu/login.aspx?direct=true&db=f6h&AN=27757513&site=eds-live

Bremer, C. A. (2008). Relationships between NWEA-map score information and teachers' efficacy beliefs for student performance on the North Dakota state assessment. ProQuest Information & Learning. *Dissertation Abstracts International Section A: Humanities and Social Sciences, 69* (3-). Retrieved from http://ezproxy.gardner-webb.edu/login?url=http://search.ebscohost.com.ezproxy.gardner-webb.edu/login.aspx?direct=true&db=psyh&AN=2008-99171-014&site=eds-live

Center for Response to Intervention. (2014). *RTI glossary of terms.* Institute American Institutes for Research and the National Center on Intensive Intervention (NCII). Washington, DC. Retrieved from http://www.rti4success.org/sites/default/files/CenterOnRTIGlossary.pdf

Clark, L. (2004) Computerized adaptive testing: Effective measurement for all students. *T.H.E. Journal, 31*(10), 18-20.

Creswell, J. W. (2007). *Qualitative inquiry and research design: Choosing among five approaches* (2nd ed.). Thousand Oaks, CA: Sage.

Creswell, J. (2013). *Research design qualitative, quantitative and mixed methods approaches* (4th ed). Thousand Oaks, CA: Sage.

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

Cronin, J., & Bowe, B. (2005). *A study of the ongoing alignment of the NWEA RIT scale with the Arizona instrument to measure standards (AIMS).* Northwest Evaluation Association. Retrieved from http://ezproxy.gardner-webb.edu/login?url=http://search.ebscohost.com.ezproxy.gardner-webb.edu/login.aspx?direct=true&db=eric&AN=ED491228&site=eds-live; http://www.nwea.org

Cronin, J., Dahlin, M., Adkins, D., Kingsbury, G. "The Proficiency Illusion," Thomas B. Fordham Institute, October 2007.

Fuchs, D., & Fuchs, L. S. (2005). Responsiveness-to-intervention: A blueprint for practitioners, policymakers, and parents. *Teaching Exceptional Children, 38,* 57-61.

Fuchs, L. S., Fuchs, D., Compton, D. L., Bryant, J. D., Hamlett, C. L., & Seethaler, P. M. (2007). Mathematics screening and progress monitoring at first grade: Implications for responsiveness to intervention. *Exceptional Children, 73,* 311-330.

George, D., & Mallery, P. (2003). SPSS for Windows step by step: A simple guide and reference. 11.0 update (4th ed.). Boston*:* Allyn & Bacon. As cited in Gliem, J & Gliem, R. Presented at the Midwest Research-to-Practice Conference in Adult, Continuing, and Community Education, The Ohio State University, Columbus, OH, October 8-10, 2003.

Gibson, S., & Dembo, M. (1984). Teacher efficacy: A construct validation. *Journal of Educational Psychology, 76,* 569-582.

Gonzales, P., Williams, T., Jocelyn, L., Roey, S., Kastberg, D., & Brenwald, S. (2008). Highlights from TIMSS 2007: Mathematics and science achievement of U.S. fourth- and eighth-grade students in an international context (NCES 2009–001 Revised). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Guskey, T. R. (1982). Differences in teachers' perceptions of personal control of positive versus negative student learning outcomes. *Contemporary Educational Psychology, 7,* 70-80.

Guskey, T., & Passaro, P. (1994). Teacher efficacy: A study of construct dimensions. *American Educational Research Journal, 31,* 627-643.

Hanson, W. E., Creswell, J. W., Plano Clark, V. L., Petska, K. S., & Creswell, J. D. (2005). Mixed methods research designs in counseling psychology. Faculty Publications, Department of Psychology. Paper 373. Retrieved from http://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1372&context=psychfacpub

Henderson, S., Petrosino, A., Guckenburg, S., & Hamilton, S. (2007). *Measuring how benchmark assessments affect student achievement. Issues & Answers. REL 2007-no. 039.* Regional Educational Laboratory Northeast & Islands. Retrieved from http://ezproxy.gardner-webb.edu/login?url=http://search.ebscohost.com.ezproxy.gardner-webb.edu/login.aspx?direct=true&db=eric&AN=ED499792&site=ehost-live

Heppen, J., Jones, W., Faria, A., Sawyer, K., Lewis, S., & Horwitz, A. (2011). *Using data to improve instruction in the great city schools: Documenting current practice. Urban data study.* Council of the Great City Schools. Retrieved from http://ezproxy.gardner-webb.edu/login?url=http://search.ebscohost.com.ezproxy.gardner-webb.edu/login.aspx?direct=true&db=eric&AN=ED536742&site=ehost-live

Heritage, M. (2007). Formative assessment: What do teachers need to know and do? *Phi Delta Kappan, 89,* 140-145.

Heritage, M. (2010). *Formative assessment: Making it happen in the classroom.* Thousand Oaks, CA: Corwin Press.

Hughes, C. A., & Dexter, D. D. (2011). Response to intervention a research-base summary. *Theory into Practice, 50,* 4-11. DOI: 10.1080/00405841.2011.534909

Jenkins, J. R. (2003, December). *Candidate measures for screening at-risk students.* Paper presented at the National Research Center on Learning Disabilities Responsiveness-to-Intervention symposium, Kansas City, MO. Retrieved May 15, 2008, from http://www.nrcld.org/symposium2003/jenkins/index.html

Jones, M. G., Jones, B. D., Hardin, B., Chapman, L., Yarbrough, T., & Davis, M. (1999). The impact of high-stakes testing on teachers and students in North Carolina. *Phi Delta Kappan, 81*(3), 199-203. Retrieved from http://ezproxy.gardner-webb.edu/login?url=http://search.ebscohost.com.ezproxy.gardner-webb.edu/login.aspx?direct=true&db=eric&AN=EJ597091&site=ehost-live

Kena, G., Aud, S., Johnson, F., Wang, X., Zhang, J., Rathbun, A., Wilkinson-Flicker, S., & Kristapovich, P. (2014). The condition of education 2014 (NCES 2014-083). Washington, DC: U.S. Department of Education, National Center for Education Statistics. Retrieved July 7, 2014 from http://nces.ed.gov/pubsearch

Kingsbury, G. G., Olson, A., Cronin, J., Hauser, C., & Houser, R. (2003). *The state of state standards: Research investigating proficiency levels in fourteen states.* Lake Oswego, NY: NWEA.

Kingsbury, G. G., & Weiss, D. J. (1983). A comparison of IRT-based adaptive mastery testing and a sequential mastery testing procedure. In D. J. Weiss (Ed.*), New horizons in testing: Latent trait test theory and computerized adaptive testing* (pp. 257-238). New York: Academic Press.

Labone, E. (2004). Teacher efficacy: Maturing the construct through research in alternative paradigms. *Teaching and Teacher Education, 20,* 341-359.

Laerd Dissertation. (n.d.). The Online Research Guide for Your Dissertation and Thesis. Retrieved from http://dissertation.laerd.com/

Laerd Statistics. (n.d.). We make statistics easy. The ultimate IBM® SPSS® guides. Retrieved from https://statistics.laerd.com/calculator

Marsh, J., & Robyn A. (2006). School and district responses to the No Child Left Behind Act, RAND Working Paper, WR-382-EDU, Santa Monica, CA: RAND Corporation. Retrieved from http://www.rand.org/ pubs/working_papers/WR382/

Masters, G. N. (2013). Reforming educational assessment: Imperatives, principles and challenges. Retrieved from http://research.acer.edu.au/aer/12

Means, B., Padilla, C., DeBarger, A., & Bakia, M. (2009). *Implementing data-informed decision making in schools: Teacher access, supports and use.* U.S. Department of Education.

National Center on Accessible Instructional Materials. (n.d.). Retrieved from http://aim.cast.org/learn/historyarchive/backgroundpapers/promise_of_udl/what_curricula#.U8hm0NxH1SU

No Child Left Behind (NCLB) Act of 2001, Pub. L. No. 107-110 & 115, Stat. 1425 (2002).

Northwest Evaluation Association. (2011a). *RIT scale norms with measures of academic progress.* Portland, OR: NWEA.

Northwest Evaluation Association. (2011b). *Technical manual for measures of academic progress and measures of academic progress for primary grades.* Portland, OR: NWEA.

Northwest Evaluation Association. (2012). *Glossary of terms.* Portland, OR: NWEA.

Northwest Evaluation Association. (2014, March). *A study of the alignment of the NWEA RIT scale with the North Carolina end of grade (EOG) testing program.* Portland, OR: NWEA.

Olson, A. (2001). Data-based change: Using assessment data to improve education. *Multimedia Schools, 8*(3), 38-43.

Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research, 66,* 543-578.

Perie, M., Marion, S., & Gong, B. (2007). *A framework for considering interim assessments.* National Center for the Improvement of Educational Assessment.

Pinchok, N., & Brandt, W. C. (2009). *Connecting formative assessment research to practice: An introductory guide for educators.* Learning Point Associates. Retrieved from http://ezproxy.gardner-webb.edu/login?url=http://search.ebscohost.com.ezproxy.gardner-webb.edu/login.aspx?direct=true&db=eric&AN=ED509943&site=ehost-live

Rose, J. S., & Medway, F. J. (1981). Measurement of teachers' beliefs in their control over student outcome. *Journal of Educational Research, 74,* 185-190.

Rossman, G. & Rallis, D. (2012). *Learning in the field: An introduction to qualitative research* (3rd ed.). Boston: Sage.

Sands, W. A., Waters, B. K., & McBride, J. R. (1997). *Computerized adaptive testing: From inquiry to operation*. Washington, DC: APA Books.

Sellman, W. S. (1988, December 14). *Memorandum for Manpower Accession Policy Steering Committee: Appointment of technical representative to a Joint-Service computerized test selection committee.* Washington, DC: Office of the Assistant Secretary of Defense.

Social Science Statistics. (n.d.). The online statistical calculator. Retrieved from http:www.socscistatistics.com

Soodak, L. C., & Podell, D. M. (1996). Teacher efficacy: Toward the understanding of a multi-faceted construct. *Teaching and Teacher Education, 12*(4), 401-411.

Tesch, R. (1990). *Qualitative research: Analysis types and software tools.* New York: Falmer.

Tschannen-Moran, M., & Woolfolk Hoy, A. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education, 17,* 783-805.

Tschannen-Moran, M., Woolfolk Hoy, A., & Hoy, W. K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research, 68,* 202-248.

UNESCO. (n.d.). Retrieved from http://en.unesco.org

United States Department of Education. (2008). Revenues and Expenditures for Public Elementary and Secondary Education: School Year 2007-08. Retrieved from https://nces.ed.gov/pubs2010/expenditures/tables/table_05.asp

Wang, S., Jiao, H., & Zhang, L. (2013). Validation of longitudinal achievement constructs of vertically scaled computerized adaptive tests: A multiple-indicator, latent-growth modelling approach. *International Journal of Quantitative Research in Education,* 1(4), 383-407.

Way, W., Twing, J., Camara, W., Sweeney, K., Lazer, S., & Mazzeo, J. (2010). *Some considerations related to the use of adaptive testing for the Common Core Assessments*. Retrieved June 11, 2010, from www.ets.org/s/commonassessments/pdf/ AdaptiveTesting.pdf

Weiner, R., & Hall, D. (2004). Accountability under No Child Left Behind. *Clearing House, 78*(1), 17-21.

Woodfield, K., & Lewis, J. (2003). Getting on board with online testing. *T.H.E Journal, 30*(6) 32-37.

Woolfolk, A. & Hoy, W. (1990). Prospective teachers' sense of efficacy and beliefs about control. *Journal of Educational Psychology, 82,* 81-91.

Woolfolk Hoy, A., Davis, H., & Pape, S. (2006). Teachers' knowledge, beliefs, and thinking. In P. A. Alexander & P. H, Winne (Eds.), *Handbook of educational psychology* (2nd ed., pp. 715-737.). Mahwah, NJ: Erlbaum.

Zimmerman, B., & Schunk, D. (Eds.). (2001). *Self-regulated learning and academic achievement: Theoretical perspectives* (2nd ed.). Mahwah, NJ: Erlbaum.

Appendix A

NWEA Normative Data Chart

2011 READING STATUS NORMS (RIT VALUES)			
Grade	Beginning-of-Year Mean	Middle-of-Year Mean	End-of-Year Mean
K	142.5	151.0	157.7
1	160.3	170.7	176.9
2	175.9	183.6	189.6
3	189.9	194.6	199.2
4	199.8	203.2	206.7
5	207.1	209.8	212.3
6	212.3	214.3	216.4
7	216.3	218.2	219.7
8	219.3	221.2	222.4
9	221.4	221.9	222.9
10	223.2	223.4	223.8
11	223.4	223.5	223.7

2011 MATHEMATICS STATUS NORMS (RIT VALUES)			
Grade	Beginning-of-Year Mean	Middle-of-Year Mean	End-of-Year Mean
K	143.7	150.7	159.1
1	162.8	172.4	179.0
2	178.2	185.5	191.3
3	192.1	198.5	203.1
4	203.8	208.7	212.5
5	212.9	217.8	221.0
6	219.6	222.8	225.6
7	225.6	228.2	230.5
8	230.2	232.8	234.5
9	233.8	234.9	236.0
10	234.2	235.5	236.6
11	236.0	237.2	238.3

Appendix B

DeCartes Continuum of Learning Instructional Statement

NWEA
Northwest Evaluation Association

DesCartes: A Continuum of Learning®

Mathematics
Goal: Number and Operations

RIT Score Range: 181 - 190
Statements Last Updated: Nov 18, 2014

Skills and Concepts to Enhance (73% Probability*) 171 - 180	Skills and Concepts to Develop (50% Probability*) 181 - 190	Skills and Concepts to Introduce (27% Probability*) 191 - 200
Understand Place Value, Counting, and Cardinality	**Understand Place Value, Counting, and Cardinality**	**Understand Place Value, Counting, and Cardinality**
• Identifies whole numbers 100 - 999 using base-10 blocks • Identifies the numerical and written name for whole numbers 21 to 100 (e.g., 62 is sixty-two, and vice versa) • Identifies the numeral and written name for whole numbers 101 to 999 (e.g., 342 is three hundred forty-two, and vice versa) • Identifies missing numbers in a series through 100 • Counts backwards from a given number (given number greater than 10) • Recognizes and generates equivalent forms for the same number using physical models for whole numbers 11 to 20 • Compares sets of objects and identifies which is equal to, more than, or less than the other (1 to 10 objects) • Compares whole numbers through 999 • Counts objects that are grouped into tens and ones • Identifies the place value and value of each digit in whole numbers through the tens place	• Identifies the numeral and written name for whole numbers 101 to 999 (e.g., 342 is three hundred forty-two, and vice versa) • Identifies the numeral and written name for whole numbers to 1000 to 9999 (e.g., 3456 is three thousand, four hundred fifty-six, and vice versa) • Identifies the numeral and written name for whole numbers 10,000 to 100,000 • Compares whole numbers through 999 • Rounds 2- and 3- digit whole numbers to the nearest ten • Rounds 3-digit whole numbers to the nearest hundred • Counts objects that are grouped into tens and ones • Identifies whole numbers under 100 given place value terms (e.g., 3 tens and 4 ones = 34) • Identifies the place value and value of each digit in whole numbers through the tens place • Identifies the place value and value of each digit in whole numbers through the hundreds place • Identifies the place value and value of each digit in whole numbers through the thousands • Identifies the place value and value of each digit in whole numbers through the hundred thousands • Compares and orders decimals to the hundredths place (same number of digits after decimal)	• Identifies whole numbers over 999 using base-10 blocks • Identifies the numeral and written name for whole numbers with a zero between digits to the ten thousands place • Identifies the numeral and written name for whole numbers 10,000 to 100,000 • Identifies the numeral and written name for whole numbers over 100,000 • Compares whole numbers to 100, using the symbols for 'less than', 'equal to', or 'greater than' (<, =, >) • Compares whole numbers through the thousands using the symbols <, >, or = • Rounds 2- and 3- digit whole numbers to the nearest ten • Rounds 3-digit whole numbers to the nearest hundred • Identifies whole numbers under 100 given place value terms (e.g., 3 tens and 4 ones = 34) • Identifies the place value and value of each digit in whole numbers through the thousands • Identifies the place value and value of each digit in whole numbers through the hundred thousands • Writes whole numbers in standard and expanded form through the hundreds • Writes whole numbers in standard and expanded form through the thousands
Number and Operations in Base Ten	**Number and Operations in Base Ten**	**Number and Operations in Base Ten**
• Uses models to calculate whole number sums through 999 • Uses strategies for addition facts (e.g., compatible numbers, counting on, doubles, neighbors, making tens) • Adds two or three 2-digit number with regrouping • Adds 1- and/or 2-digit numbers with sums under 100 • Adds 3-digit numbers with no regrouping • Adds 3-digit numbers, with regrouping, with sums under 1000 • Subtracts a 2-digit number from a 2-digit number, with no regrouping • Subtracts 2- and/or 3-digit numbers with no regrouping	• Adds two or three 2-digit number with regrouping • Adds 3-digit numbers, with regrouping, with sums under 1000 • Performs mental computation with 2, 3, or 4 addends • Adds two 3- and/or 4-digit numbers, with regrouping, with sums over 1000 • Adds multiple-digit numbers, with regrouping, with sums over 1000 • Uses models to calculate differences through 100 (whole numbers) • Subtracts a 2-digit number from a 2-digit number, with regrouping • Uses strategies for sums and differences with 2-digit numbers (e.g., decomposing, compatible, compensation, partial sums, counting on) • Subtracts 2- and/or 3-digit numbers with no regrouping • Subtracts 3- or 4-digit numbers with regrouping • Performs mental subtraction with numbers under 1000 • Subtracts multiple-digit numbers with no regrouping • Multiplies a 2-digit number by a 1-digit number with regrouping • Multiplies a 2-digit number by a 2-digit number with no regrouping	• Uses rounding to estimate answers to addition and subtraction problems (whole numbers only) • Adds two 3- and/or 4-digit numbers, with regrouping, with sums over 1000 • Adds multiple-digit numbers, with regrouping, with sums over 1000 • Adds multiple-digit numbers with sums under 1000 • Subtracts 1-digit number from a 2-digit number with regrouping • Subtracts a 2-digit number from a 2-digit number, with regrouping • Uses strategies for sums and differences with 2-digit numbers (e.g., decomposing, compatible, compensation, partial sums, counting on) • Subtracts a 2-digit number from a 3-digit number with a single regrouping • Subtracts 3- or 4-digit numbers with regrouping • Performs mental subtraction with numbers under 1000 • Subtracts multiple-digit numbers with no regrouping • Multiplies a 2- or 3-digit number by a 1-digit number with no regrouping • Multiplies a 2-digit number by a 1-digit number with regrouping

Appendix C

Teacher Beliefs Survey

Teacher Beliefs

Thank you for taking time to respond to this survey. This information is being used as part of a doctoral dissertation study. The information you provide is anonymous. School Name is being used only by the researcher and will be coded as School A, B, etc. for publication. Please respond in a manner that best describes your opinion.

* Required

I am a teacher at *

Please indicate the elementary school in which you currently teach

I primarily teach *

- Kindergarten
- 1st grade
- 2nd grade
- 3rd grade
- 4th grade
- 5th grade

I have been teaching for *

- 15+ years
- 10-14 years
- 5-9 years
- 1-4 years
- Less than 1 full school year

I have been teaching in a school participating in Measures of Academic Progress (MAP) assessments for *

(including this school year)

- More than 3 years
- 3 years
- 2 years
- 1 year

Rating Scale Items

The following rating scale has two types of items--level of agreement and frequency. Please rate your responses according the following descriptors:
1=Strongly Disagree; 2=Disagree; 3=Neutral; 4=Agree; 5=Strongly Agree
1=Never; 2=Rarely; 3=Sometimes; 4=Often; 5=Frequently

1) I know how to access all of the MAP information about my students. *

	1	2	3	4	5	
Strongly Disagree	○	○	○	○	○	Strongly Agree

2) I understand how to use the DesCartes Continuum of Learning. *

	1	2	3	4	5	
Strongly Disagree	○	○	○	○	○	Strongly Agree

3) I understand how data from MAP can be used to make classroom instructional decisions. *

	1	2	3	4	5	
Strongly Disagree	○	○	○	○	○	Strongly Agree

4) I understand how student growth is measured by MAP. *

	1	2	3	4	5	
Strongly Disagree	○	○	○	○	○	Strongly Agree

5) I understand how RIT bands measure student understanding of concepts. *

	1	2	3	4	5	
Strongly Disagree	○	○	○	○	○	Strongly Agree

6) I can explain my students' academic abilities using information from MAP. *

	1	2	3	4	5	
Strongly Disagree	○	○	○	○	○	Strongly Agree

7) I use information from MAP reports to make instructional decisions. *

1 2 3 4 5

Never ○ ○ ○ ○ ○ Frequently

8) I use the DesCartes Continuum of Learning to make instructional decisions. *

1 2 3 4 5

Never ○ ○ ○ ○ ○ Frequently

9) I talk with my PLC or other teachers about MAP data. *

1 2 3 4 5

Never ○ ○ ○ ○ ○ Frequently

10) I communicate with my students about goals related to MAP. *

1 2 3 4 5

Never ○ ○ ○ ○ ○ Frequently

11) I use my students' MAP data to make instructional decisions. *

1 2 3 4 5

Never ○ ○ ○ ○ ○ Frequently

12) I use each students' DesCartes Continuum of Learning to make instructional decisions. *

1 2 3 4 5

Never ○ ○ ○ ○ ○ Frequently

13) My students' MAP data is an accurate reflection of their academic knowledge. *

1 2 3 4 5

Strongly Disagree ○ ○ ○ ○ ○ Strongly Agree

14) I set growth goals for myself related to MAP. *

1 2 3 4 5

Strongly Disagree ○ ○ ○ ○ ○ Strongly Agree

15) My students' MAP growth is an accurate reflection of their learning in my classroom. *

1 2 3 4 5

Strongly Disagree ○ ○ ○ ○ ○ Strongly Agree

16) My students' MAP performance is comparable to other sources of data about them. *

1 2 3 4 5

Strongly Disagree ○ ○ ○ ○ ○ Strongly Agree

17) My time spent understanding MAP data is well-spent. *

1 2 3 4 5

Strongly Disagree ○ ○ ○ ○ ○ Strongly Agree

18) My work with MAP data informs my lesson planning. *

1 2 3 4 5

Appendix D

Survey Items by Construct

Item	Item Number	Construct
I know how to access all of the MAP information about my students.	1	Teacher Capacity
I understand how to use the DesCartes Continuum of Learning.	2	Teacher Capacity
I understand how data from MAP can be used to make classroom instructional decisions.	3	Teacher Capacity
I understand how student growth is measured by MAP.	4	Teacher Capacity
I understand how RIT bands measure student understanding of concepts.	5	Teacher Capacity
I can explain my students' academic abilities based on information from MAP.	6	Teacher Capacity
I use information from MAP reports to make instructional decisions.	7	Teacher Attitude
I use the DesCartes Continuum of learning to make instructional decisions.	8	Teacher Attitude
I talk with my PLC or other teachers about my MAP data.	9	Teacher Attitude
I communicate with my students about goals related to MAP.	10	Teacher Attitude
I use my students' MAP data to make instructional decisions.	11	Teacher Attitude
I use each students' Descartes Continuum of Learning to make instructional decisions.	12	Teacher Attitude
My students' MAP data is an accurate reflection of their academic knowledge.	13	Teacher Attitude
I set growth goals for myself related to MAP.	14	Teacher Attitude
My students' MAP growth is an accurate reflection of their learning in my classroom.	15	Teacher Attitude
My students' MAP data is comparable to other sources of data about them.	16	Teacher Attitude
My time spent understanding MAP data is well-spent.	17	Teacher Attitude
My work with MAP data informs my lesson planning.	18	Teacher Attitude

Appendix E

Permission to Use TSES

Elizabeth Mitcham,

You have my permission to adapt and use the Teacher Sense of Efficacy Scale (formerly called the Ohio State Teacher Sense of Efficacy Scale) that I developed with Anita Woolfolk Hoy in your dissertation research. You can find a copy of the measure and scoring directions on my web site at http://wmpeople.wm.edu/site/page/mxtsch. Please use the following as the proper citation (even though the earlier name was used in that article):

Tschannen-Moran, M & Hoy, A. W. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education, 17*, 783-805.

I will also attach directions you can follow to access my password protected web site, where you can find the supporting references for this measure as well as other articles I have written on this and related topics. I would love to receive a brief summary of your results.

All the best,

Megan Tschannen-Moran
The College of William & Mary
School of Education
PO Box 8795
Williamsburg, VA 23187-8795
757-221-2187
http://wmpeople.wm.edu/site/page/mxtsch

Appendix F

Focus Group Protocol School B

Facilitator: Hello and welcome. My name is Sabrina Fordham, and I will be conducting the focus group today. Elizabeth wanted me to thank you for your participation in this session. The session will be no more than 45 minutes in length. Please help yourself to refreshments.
Facilitator: First, I need you to sign the consent forms. We will be audio-recording this session for transcription as part of the process. Please do not use you names or any other teachers' names during the taping. You can reference "the principal" "our school" or "4th grade teachers."

Questions about the survey data:

Looking at the Capacity and Attitude Analysis, this table tells you the average score Capacity score out of 30 and the average percentage of Teacher Attitude out of 100% by grade level. Take some time to look at this data.

1) What are the initial impressions about this data?

2) What factors about how you school uses MAP could have attributed to this data?

3) The overall attitude score for teacher perceptions about the MAP assessment, MAP data and usage of MAP data at your school was 41.4%. This was statistically lower than the other schools in the study. Can you think of any reasons the data may reflect that?

**Follow Up: Capacity data indicates that the teachers in your school believe that they are capable of accessing and utilizing the data to effectively plan instruction, what are your thoughts on this compared to the overall attitude score?

Looking at the Item Analysis Data Table, this table tells you the average score on a scale of 1-5 for items related specifically to personal feelings about MAP and its data. Data is provided for your school and the average for all the schools in the study.

1) Item 13 is a measure of whether teachers believe MAP scores are accurate reflections of student abilities and measures of learning in their classrooms. What are your thoughts about the scores for your school? Aggregate data? What surprises you about this data?

2) Item 15 is a measure of how strongly teachers' believed that their students MAP growth was a reflection of the learning in their classrooms. How might this be explained?

3) Item 16 measured teachers' perceptions of how comparable MAP data was to other data sources about their students. What are your thoughts?

Final Questions:

4) Is there anything else anyone would like to say about the results of the survey presented today?

5) Is there anything else anyone would like to share about their beliefs about this topic?

Appendix G

Focus Group Protocol School D

Facilitator: Hello and welcome. My name is Sabrina Fordham, and I will be conducting the focus group today. Elizabeth wanted me to thank you for your participation in this session. The session will be no more than 45 minutes in length. Please help yourself to refreshments.
Facilitator: First, I need you to sign the consent forms. We will be audio-recording this session for transcription as part of the process. Please do not use you names or any other teachers' names during the taping. You can reference "the principal" "our school" or "4th grade teachers."

Questions about the survey data:

Looking at the Capacity and Attitude Analysis, this table tells you the average score Capacity score out of 30 and the average percentage of Teacher Attitude out of 100% by grade level. Take some time to look at this data.

1) What are the initial impressions about this data?

2) What factors about how you school uses MAP could have attributed to this data?

3) The average attitude score for 1st grade was statistically considered to show that teachers had "highly unfavorable" attitudes towards the data from MAP assessments. What may attribute to this?

***If they ask, more than 50% of the teachers in 1st grade responded to the survey, so this is considered to be the primary view of the grade level.*

**FOLLOW UP 1: Is this data from 1st grade surprising to you? Are there other data sources that you aware of that suggest a similar thing?
**FOLLOW UP 2: Capacity data indicates that the teachers in your school believe that they are capable of accessing and utilizing the data from MAP to effectively plan instruction, what are your thoughts on this compared to the overall attitude score?

Looking at the Item Analysis Data Table, this table tells you the average score on a scale of 1-5 for items related specifically to personal feelings about MAP and its data. Data is provided for your school and the average for all the schools in the study.

1) Item 13 is a measure of whether teachers believe MAP scores are accurate reflections of student abilities and academic knowledge. What are your thoughts about the scores for your school? Aggregate data? What surprises you about this data?

2) Item 15 is a measure of how strongly teachers' believed that their students MAP growth was a reflection of the learning in their classrooms. How might this be explained?

3) Item 16 measured teachers' perceptions of how comparable MAP data was to other data sources about their students. What are your thoughts?

Final Questions:

4) Is there anything else anyone would like to say about the results of the survey presented today?

5) Is there anything else anyone would like to share about their beliefs about this topic?

Appendix H

Focus Group Protocol School A

Facilitator: Hello and welcome. My name is Sabrina Fordham, and I will be conducting the focus group today. Elizabeth wanted me to thank you for your participation in this session. The session will be no more than 45 minutes in length. Please help yourself to refreshments.
Facilitator: First, I need you to sign the consent forms. We will be audio-recording this session for transcription as part of the process. Please do not use you names or any other teachers' names during the taping. You can reference "the principal" "our school" or "4th grade teachers."

Questions about the survey data:

Looking at the Capacity and Attitude Analysis, this table tells you the average score Capacity score out of 30 and the average percentage of Teacher Attitude out of 100% by grade level. Take some time to look at this data.

1) What are the initial impressions about this data?

2) What factors about how you school uses MAP could have attributed to this data?

3) The overall attitude score for teacher perceptions about the MAP assessment, MAP data and usage of MAP data at your school was 41.4%. This was statistically lower than the other schools in the study. Can you think of any reasons the data may reflect that?

**Follow Up: Capacity data indicates that the teachers in your school believe that they are capable of accessing and utilizing the data to effectively plan instruction, what are your thoughts on this compared to the overall attitude score?

Looking at the Item Analysis Data Table, this table tells you the average score on a scale of 1-5 for items related specifically to personal feelings about MAP and its data. Data is provided for your school and the average for all the schools in the study.

1) Item 13 is a measure of whether teachers believe MAP scores are accurate reflections of student abilities and measures of learning in their classrooms. What are your thoughts about the scores for your school? Aggregate data? What surprises you about this data?

2) Item 15 is a measure of how strongly teachers' believed that their students MAP growth was a reflection of the learning in their classrooms. How might this be explained?

3) Item 16 measured teachers' perceptions of how comparable MAP data was to other data sources about their students. What are your thoughts?

Final Questions:

4) Is there anything else anyone would like to say about the results of the survey presented today?

5) Is there anything else anyone would like to share about their beliefs about this topic?

Appendix I

MAP Grade Report Summary

NWEA
Northwest Evaluation Association

Grade Report

Grade 2

Term:	Fall 2010-2011
District:	NWEA Sample District 3
School:	Three Sisters Elementary School
Grouping:	None
Small Group Display:	No

Reading

Reading Survey w/ Goals 2-5 CO V2.1

Summary	
Total Students With Valid Growth Test Scores	137
1 Mean RIT	178.4
2 Standard Deviation	14.9
3 District Grade Level Mean RIT	175.6
4 Students At or Above District Grade Level Mean RIT	73
5 Norm Grade Level Mean RIT	175.9
6 Students At or Above Norm Grade Level Mean RIT	73

7 Overall Performance	Lo %ile < 21		LoAvg %ile 21-40		Avg %ile 41-60		HiAvg %ile 61-80		Hi %ile > 80		9 Mean RIT (+/- Smp Err)	10 Std Dev
	count	%	count	%	count	%	count	%	count	%		
Reading Survey w/ Goals 2-5 CO V2.1	29	21%	21	15%	26	19%	22	16%	39	28%	177-178-180	14.9
8 Goal Area												
Students Read and Understand Variety of Material	30	22%	20	15%	28	20%	22	16%	37	27%	176-178-179	16.7
Students Apply Thinking Skills to Their Reading	29	21%	26	19%	17	12%	28	20%	37	27%	177-179-180	16.7
Students Locate, Select, and Use Information	18	13%	37	27%	30	22%	17	12%	35	26%	177-179-180	14.8
Students Read and Recognize Literature	28	20%	25	18%	17	12%	27	20%	40	29%	178-179-181	16.7

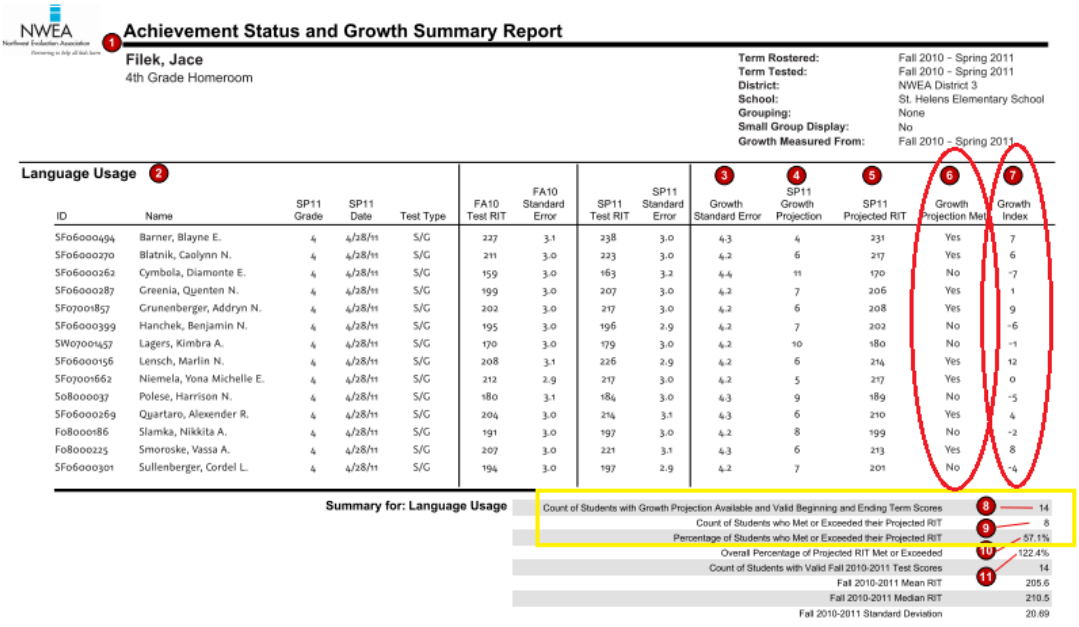
Explanatory Notes

Tests shown in gray are excluded from summary statistics. Either the test occurred outside the testing window for a term, had an invalid score, or was a repeat test for a student within a term.
Due to statistical unreliability, summary data for groups of less than 10 are not shown.
* This data is not available for reporting. Please refer to help and documentation for more information.

Appendix J

MAP Achievement Status and Growth Report Summary

MAP Reports Reference



NWEA

Achievement Status and Growth Summary Report (1)

Filek, Jace
4th Grade Homeroom

Term Rostered:	Fall 2010 – Spring 2011
Term Tested:	Fall 2010 – Spring 2011
District:	NWEA District 3
School:	St. Helens Elementary School
Grouping:	None
Small Group Display:	No
Growth Measured From:	Fall 2010 – Spring 2011

Language Usage (2)

ID	Name	SP11 Grade	SP11 Date	Test Type	FA10 Test RIT	FA10 Standard Error	SP11 Test RIT	SP11 Standard Error	(3) Growth Standard Error	(4) SP11 Growth Projection	(5) SP11 Projected RIT	(6) Growth Projection Met	(7) Growth Index
SF06000494	Barner, Blayne E.	4	4/28/11	S/G	227	3.1	238	3.0	4.3	4	231	Yes	7
SF06000270	Blatnik, Caolynn N.	4	4/28/11	S/G	211	3.0	223	3.0	4.2	6	217	Yes	6
SF06000262	Cymbola, Diamonte E.	4	4/28/11	S/G	159	3.0	163	3.2	4.4	11	170	No	-7
SF06000287	Greenia, Quenten N.	4	4/28/11	S/G	199	3.0	207	3.0	4.2	7	206	Yes	1
SF07001857	Grunenberger, Addryn N.	4	4/28/11	S/G	202	3.0	217	3.0	4.2	6	208	Yes	9
SF06000399	Hanchek, Benjamin N.	4	4/28/11	S/G	195	3.0	196	2.9	4.2	7	202	No	-6
SW07001457	Lagers, Kimbra A.	4	4/28/11	S/G	170	3.0	179	3.0	4.2	10	180	No	-1
SF06000156	Lensch, Marlin N.	4	4/28/11	S/G	208	3.1	226	2.9	4.2	6	214	Yes	12
SF07001662	Niemela, Yona Michelle E.	4	4/28/11	S/G	212	2.9	217	3.0	4.2	5	217	Yes	0
S08000037	Polese, Harrison N.	4	4/28/11	S/G	180	3.1	184	3.0	4.3	9	189	No	-5
SF06000269	Quartaro, Alexender R.	4	4/28/11	S/G	204	3.0	214	3.1	4.3	6	210	Yes	4
F08000186	Slamka, Nikkita A.	4	4/28/11	S/G	191	3.0	197	3.0	4.2	8	199	No	-2
F08000225	Smoroske, Vassa A.	4	4/28/11	S/G	207	3.0	221	3.1	4.3	6	213	Yes	8
SF06000301	Sullenberger, Cordel L.	4	4/28/11	S/G	194	3.0	197	2.9	4.2	7	201	No	-4

Summary for: Language Usage

(8) Count of Students with Growth Projection Available and Valid Beginning and Ending Term Scores	14
(9) Count of Students who Met or Exceeded their Projected RIT	8
Percentage of Students who Met or Exceeded their Projected RIT	57.1%
(10) Overall Percentage of Projected RIT Met or Exceeded	122.4%
(11) Count of Students with Valid Fall 2010-2011 Test Scores	14
Fall 2010-2011 Mean RIT	205.6
Fall 2010-2011 Median RIT	210.5
Fall 2010-2011 Standard Deviation	20.69