2014

Is the college-ready teaching framework related to student achievement?

Narciso Aguda

Pepperdine University

Graduate School of Education and Psychology

IS THE COLLEGE-READY TEACHING FRAMEWORK RELATED TO STUDENT ACHIEVEMENT?

A dissertation submitted in partial satisfaction

of the requirements for the degree of

Doctor of Education in Learning Technologies

by

Narciso Aguda

July, 2014

Farzin Madjidi, Ed.D. – Dissertation Chairperson

This dissertation, written by

Narciso Aguda

under the guidance of a Faculty Committee and approved by its members, has been submitted to and accepted by the Graduate Faculty in partial fulfillment of the requirements for the degree of

DOCTOR OF EDUCATION

Doctoral Committee:

Farzin Madjidi, Ed.D., Chairperson

Paul Sparks, Ed.D.

Kevin Keelen, Ed.D.

TABLE OF CONTENTS

Page

LIST OF TABLES vi

LIST OF FIGURES viii

DEDICATION ix

ACKNOWLEDGMENTS x

VITA xii

ABSTRACT xiii

Chapter 1: Introduction and Background of the Study 1

- Background of the Study 1
- Statement of the Problem 4
- Theoretical Framework 6
- Purpose of the Study 7
- Research Questions 7
- Delimitations and Limitations 8
- Definition of Terms 9

Chapter 2: Review of Literature 12

- Background 12
- Defining Teacher Effectiveness 15
- Teacher Evaluation 20
- Summary 35

Chapter 3: Methodology 37

- Research Design 37
- Description and Selection of the Subject Sample 38
- Data Collection 40
- Instrumentation 44
- Restatement of Research Questions 49
- Variables in the Study and Their Levels of Measurement 50
- Statistical Analysis 50

Chapter 4: Results 52

- Description of the Sample 52
- Research Question #1 52
- Research Question #2 55

Research Question #3 58
Research Question #4 71
Research Question #5 74
Research Question #6 85

Chapter 5: Conclusions and Implications 89

Restatement of Purpose and Research Questions 89
Summary of Key Findings 90
Discussion of Results 91
Implications of the Study 93
Limitations 95
Recommendations for Future Research 97
Summary 98

REFERENCES 99

APPENDIX A: IRB Approval 106

APPENDIX B: The College-Ready Teaching Framework 107

LIST OF TABLES

Page

Table 1. Teacher Effectiveness Score Breakdowns 45

Table 2. Student Growth Percentiles: What They Mean 48

Table 3. Variable Table 50

Table 4. Correlation Matrix for Research Question #1 53

Table 5. Correlation Matrix for Research Question #1 Disaggregated (ELA Only) 54

Table 6. Correlation Matrix for Research Question #1 Disaggregated (History Only) 54

Table 7. Correlation Matrix for Research Question #1 Disaggregated (Math Only) 55

Table 8. Correlation Matrix for Research Question #1 Disaggregated (Science Only) 55

Table 9. Correlation Matrix for Research Question #2 56

Table 10. Correlation Matrix for Research Question #2 Disaggregated (Science Only) 57

Table 11. Correlation Matrix for Research Question #2 Disaggregated (Math Only) 57

Table 12. Correlation Matrix for Research Question #2 Disaggregated (History Only) 58

Table 13. Correlation Matrix for Research Question #2 Disaggregated (ELA Only) 58

Table 14. Correlation Matrix for Research Question #3 - All Subjects 59

Table 15. Correlation Matrix for Research Question #3 - Disaggregated (Science Only) 60

Table 16. Correlation Matrix for Research Question #3 - Disaggregated (Math Only) 61

Table 17. Correlation Matrix for Research Question #3 - Disaggregated (History Only) 62

Table 18. Correlation Matrix for Research Question #3 Disaggregated (ELA Only) 62

Table 19. Multiple Regression Summary Table (All Subjects) – Showing Final Iterations 64

Table 20. Multiple Regression Summary Table (Disaggregated for ELA Only) – Showing Final Iterations 67

Table 21. Multiple Regression Summary Table (Disaggregated for History Only) – Showing Final Iterations 68

Table 22. Multiple Regression Summary Table (Disaggregated for Math Only) – Showing Final Iterations 69

Table 23. Multiple Regression Summary Table (Disaggregated for Science Only) – Showing Final Iterations 70

Table 24. Factor Loadings and Communalities Based on Factor Analysis with Varimax Rotation for 29 Items from Green Dot Public School's CRTF 72

Table 25. Correlation Matrix for New Factors (All Subjects) 74

Table 26. Correlation Matrix for Factors (ELA Only) 75

Table 27. Correlation Matrix for Factors (History Only) 76

Table 28. Correlation Matrix for Factors (Math Only) 76

Table 29. Correlation Matrix for Factors (Science Only) 77

Table 30. Multiple Regression Summary Table for New Factors (All Subjects) – Showing Final Iterations 78

Table 31. Multiple Regression Summary Table for New Factors (ELA Only) – Showing Final Iterations 81

Table 32. Multiple Regression Summary Table for New Factors (History Only) – Showing Final Iterations 82

Table 33. Multiple Regression Summary Table for New Factors (Math Only) – Showing Final Iterations 83

Table 34. Multiple Regression Summary Table for New Factors (Science Only) – Showing Final Iterations 84

Table 35. Step-wise Regression Summary Chart (All Subjects) 86

Table 36. Step-wise Regression Summary Chart (ELA Only) 86

Table 37. Step-wise Regression Summary Chart (History Only) 87

Table 38. Step-wise Regression Summary Chart (Math Only) 87

Table 39. Step-wise Regression Summary Chart (Science Only) 88

LIST OF FIGURES

Page

Figure 1. Sample organization of the CRTF 45

Figure 2. Results of factor analysis showing four new factors compared to original CRTF Domains. 73

DEDICATION

My educational career up to and including this dissertation is dedicated to my late father, Narciso P. Aguda, Sr., who passed away in December of 2007 after a short battle with Leukemia. For as long as I can remember, he valued education and the pursuit of knowledge. He truly believed that education and hard work could open doors to numerous opportunities in life and career. My father embodied these values by investing in the lives of his own brothers and sisters as well as instilling them into my sister's life and mine. I've carried these values to this terminal degree: a degree that I'm more than sure has made him proud. This one's for you, dad. I miss you and love you.

ACKNOWLEDGMENTS

So many fantastic people helped me through this journey that to thank each individual specifically would increase the page count of this work exponentially. To everyone who has said a prayer, offered encouragement, given me a high-five, and/or bought me a can of Coke Zero, I offer my sincerest thanks and appreciation. I couldn't have done it without a support system of friends and family like you.

To my dissertation committee:

- Dr. Farzin Madjidi, Chairperson. Thank you immensely for your help with this work. My being green to writing a dissertation did not seem to faze you. You helped me set high expectations for completing the work with quality and excellence. In addition, though I recall that I did well in your inferential statistics course several semesters back, you were patient to help me through the more advanced statistics. Thank you so very much for the late nights crunching the data while sharing your food, Kreation cold-pressed juice, and Smith, your pit bull puppy.
- Dr. Paul Sparks. Thank you so much for serving on my committee and providing invaluable input during the process. Your positivity, passion, and energy have always been something I've admired in you as an instructor. These values proved to be instrumental throughout this process.
- Dr. Kevin Keelen. I don't even know where to begin. Thanks so much for serving on the committee and for being so supportive through the process. Your attention to detail and passion for data and statistics were some of the reasons I became interested in a quantitative analysis. Your role at Green Dot has made this process seamless and I am forever grateful.

To Green Dot Public Schools and Knowledge Management (KM):

- I am thankful for my KM family of analysts and knowledge managers that have offered so much love and support to me through the process.
- Special thanks to my boss, Samantha Mita, who took the time to listen to me ramble about the process during our one-on-ones and offered her help in spot checking data.
- To Julia Fisher of Green Dot Teacher Effectiveness fame, who, over a beer at some downtown LA bar in what seems like a forever-and-a-half ago, helped me shape this study in its initial stages. Thanks for your support.

To my sister, Ferlay, thanks for your love and for always believing I could reach my dreams. The challenge of writing has always been your strong suit, and I've taken cues from you throughout this journey. Thanks.

To my mom, I love you. You've always encouraged me to be the best and attain the highest. You are a constant reminder of daddy. Thanks for always praying for me.

To my two cuddly beagles, Jack and Tiger, who were my buddies during those long writing nights and sat patiently through my presentation rehearsals.

Finally, to Rocío Garcia, I really think you could accurately present the findings to this study since I've spent hours talking to you about it! You have been my strength and support throughout this entire journey. You've always been there to listen, encourage, and challenge me. I couldn't have gotten here without you. Your selflessness, sacrifice, and love helped to center my eyes on the prize. I love you. #jayroventures.

VITA

EDUCATION

2014	Doctorate of Education in Learning Technologies Pepperdine University
2006	Master of Education University of California Los Angeles
2003	Bachelor of Science in Physiological Science University of California Los Angeles

PROFESSIONAL HISTORY

2010 – Present	Instructional Technology Coordinator Green Dot Public Schools
2006 – 2010	Technology Teacher Ánimo Justice Charter High School Green Dot Public Schools
2005 – 2006	Science Instructor Santee Education Complex Los Angeles Unified School District
2004 – 2005	Science Instructor (Student) Alain LeRoy Locke High School Los Angeles Unified School District
1999 – 2008	Youth Pastor Faith Tabernacle Church

ABSTRACT

This dissertation examined the College-Ready Teaching Framework (CRTF), a multiple measure teacher effectiveness rubric created by Green Dot Public Schools. The purpose of the dissertation was to determine whether or not the CRTF as a whole or in part could account for differences in student outcomes (California Standards Test [CST] scores, student growth percentile [SGP], and grade point average [GPA]). The study included teachers and students at Green Dot during the 2012-2013 school year. Correlational analyses were used to determine if there was a relationship between student achievement outcomes and the CRTF. Factor analysis was used to discover other Factors in addition to the CRTF's original five Domains. Multiple regression and step-wise regression were employed to determine if a combination of indicators, Domains, or Factors could predict student scores. The results of the findings showed that overall there were no relationships between Teacher Effectiveness Score (TES), Teacher Observation Score (TObs), and student outcome metrics (SGP, CST, and GPA). Disaggregating the dataset for math, science, and history separately, however, moderate relationships emerged between TES, TObs, SGP, and CST. Four additional Factors emerged from factor analysis that were similar to the original theoretical Domains created by CRTF designers; however. neither the original Domains nor the additional Factors were related to student outcomes. Finally, no regression model was found to hold any practical significance as no combination of indicators, Domains, or Factors accounted for more than 19.5% of the variation in student outcomes. The findings of this study are largely consistent with similar studies in the research literature where correlation analysis has been promising, yet inconsistent. The results of this study represent the addition of the CRTF to the research literature. Future research on the study of the effect of professional development and the impact of various weights of the CRTF composite score are recommended.

Chapter 1: Introduction and Background of the Study

Background of the Study

On the verge of another reauthorization of the Elementary and Secondary Education Act (ESEA), the climate of education is again changing. President Lyndon B. Johnson enacted the original act in 1965 to address inequalities in access to education and bridge the achievement gap. The law's goal was to provide every student with equal access to education in an effort to bridge the achievement gap. Nearly 4 decades later, the George W. Bush administration's reauthorization of ESEA, also known as the No Child Left Behind Act (NCLB, 2002), developed a highly rigid system of accountability that placed the burden heavily on Title 1 schools to meet Adequate Yearly Progress (AYP) in order to receive funding. The act also required states to ensure that all teachers are *highly qualified*, possessing the certifications and education qualifications to teach to his/her subject matter. NCLB has ambitious student achievement ends: requiring *all* students to achieve proficiency by 2014. The federal government's stringent requirements promptly changed the overall climate: namely, setting high standards and expectations for students via meeting yearly goals. However, Darling-Hammond (2004) attributed the lingering achievement gap, especially for students in low-income areas, to divergent teacher quality. Today, however, an increased interest in teacher accountability pervades the educational landscape.

Highly qualified teachers and teacher quality. A *highly qualified* teacher was once considered to be someone possessing the proper paper credentials and qualifications. However, today the effectiveness of teachers is measured by their students' achievements. In 2013, the Measures of Effective Teaching (MET) study concluded that teachers who have been previously identified as effective proved their effectiveness by their ability to help their students learn.

Darling-Hammond (2000) and Wright, Horn, and Sanders (1997) effectively linked the determining factor, in regard to student success and improvement, to students' classroom teachers. Hanushek (1992) revealed that students who were taught by a good teacher exhibited a difference of a full grade level of achievement in 1 school year in comparison to students who had had a bad teacher. Additionally, Sanders and Rivers (1996) found that, over time, student gains from having an effective teacher, peaked at 50 percentile points. Furthermore, the data indicated that cumulative effects of multiple effective teachers could be aggregated over a 3-year period.

Measuring teacher quality through value add and growth measurements. One measure of teacher quality, according to Goe (2007), is a teacher's overall effectiveness. Today's educational policy has welcomed statistical modeling to determine a teacher's effectiveness either as a calculation of the teacher's value-add or a measure of his/her students' growth. Value added models (VAMs), already used in business econometrics and biological sciences to model change, have also emerged as a valid measure of teacher effectiveness because they are able to demonstrate student gains throughout the achievement spectrum (Slaughter, 2008). Studies of using VAMs such as those by Hanushek (1992), Sanders and Rivers (1996), and Wright et al. (1997) strongly suggested that the effect of teachers may indeed be larger than socioeconomic influences and student background, which were originally thought to be major contributors to lower student achievement (McCaffrey, Lockewood, Koretz, & Hamilton, 2003). Because of such claims, the use of VAMs as an indicator of teacher effectiveness has sparked political debate surrounding teacher evaluation in recent years. In growth models, teacher quality has been measured by how much growth an average student in a teacher's class makes in one year, typically comparing a student's growth from one year to the next on a standardized test. A student of an effective teacher would grow significantly more than if he/she were a student of a less effective teacher (Hanushek, 1992; Sanders & Rivers, 1996). Student growth percentiles offer the promise of disaggregating the effects of non-educational factors such as socio-economic status and race.

Racing to the top of teacher effectiveness programs. With a deeper emphasis on a teacher's practice as a means to yield better student achievement, schools and districts can actionably implement these teacher practice measures to hold teachers accountable for student achievement while also developing robust professional development programs. Today, to address the concerns of high turnover rates in certain high-risk public schools, some districts have formed talent management and human capital departments to ensure the recruiting, development, and retention of highly effective teachers (Odden & Kelly, 2008). By using formal observation data collected via the rubric, an actionable plan can be created. A teacher's score on rubric indicators can help administrators provide targeted professional development to a struggling teacher. Taken in aggregate, score trends are also useful for district-level instructional coaches to support overall professional development during in-service training days.

The federal government realizes the power of teacher effectiveness programs as well: offering states and districts the opportunity to develop them in exchange for flexibility from the strict school-accountability measures of NCLB. Their contribution so far has been substantial, offering over $4.35 billion in competitive grants and federal policies such as Race to the Top and the Teacher Incentive Fund to spark innovation in teacher effectiveness systems. (U.S. Department of Education, 2009).

Many states and districts have answered this call by submitting applications to such programs. Green Dot Public Schools (Green Dot), a charter management organization (CMO) in

Los Angeles, California, is one such district. Green Dot is at the forefront of the work in teacher evaluation systems, committed to ensuring that a highly effective teacher leads each classroom. They—along with other CMOs such as Alliance, PUC, and Aspire Public Schools—have formed a coalition known as The College Ready Promise (TCRP), funded by the Bill and Melinda Gates Foundation. The coalition aims to develop and implement a framework for teacher and administrator effectiveness reforms. To accomplish this, TCRP developed the College Ready Teaching Framework (CRTF), a multiple-measure teacher evaluation system. Loosely based on Danielson's (2011) *Framework for Teaching*, the CRTF incorporates a teacher's observation score (Domains 1-4), individual student growth percentile (SGP), school-level SGP, and various stakeholder feedback to determine a teacher's overall TES. The TES is a teacher's yearly official formal evaluation.

Statement of the Problem

A wealth of research has documented the positive effect certain teacher characteristics and teaching practice can have on student outcomes. Research has shown that many of the differences in student learning can be largely attributed to *teacher effects* and *teacher practices* or what a teacher actually brings to the classroom (Goe, 2007; Rivkin, Hanushek, & Kain, 2005). This is especially promising because the improvement of teaching practice is the primary goal of teacher professional development. However, no consensus exists on what sets or combinations of teacher effects and behaviors (in the form of teacher effectiveness rubrics or otherwise) consistently produce desired student outcomes (Goe, 2007). Danielson's (2011) *A Framework for Teaching,* on which the CRTF is based, suggests five Domains comprising the teacher effects that characterize good teaching. The framework is a popular rubric for determining teacher quality through 22 components of teaching practice. The tool helps administrators, schools, and districts evaluate and develop their teachers. The framework defines the 22 competencies of excellent teaching through proficiency in four Domains: planning and preparation, classroom environment, instruction, and professional responsibilities. Green Dot's version of the framework includes data-driven planning and assessing of student learning (Domain 1), classroom learning environment (Domain 2), instruction (Domain 3), developing professional practice (Domain 4), and developing partnerships with family and community (Domain 5).

Though the CRTF was developed over 3 years of research and stakeholder feedback, only internal statistical analysis has been conducted on the data regarding its impact on student outcomes. The analysis seems to be consistent with current research showing no strong positive relationships between overall teacher evaluation score and student outcomes (Borman & Kimball, 2005; Cohen & Hill, 1998; Goe, 2007). The lack of convincing evidence in the research literature linking teacher practices to student outcomes is problematic, as no rationale exists to continue to develop these kinds of systems within the organization. The significance of the findings provides Green Dot with a rationale to continue to develop or augment the current rubric and to focus on professional development related to indicators, Domains, or factors shown to be related to student achievement. Furthermore, the five Domains measured in the CRTF were identified by the designers of the instrument during development. There have been no numerical analyses of the CRTF data that have revealed other constructs (other than the original Domains identified) nor has there been analyses on whether or not those constructs have an influence on student outcomes. Therefore, this study explored whether the CRTF as a whole or in part can account for differences in student outcomes (e.g., grade point average [GPA], standardized test scores, and student growth). In addition, the proposed study sought to uncover any additional factors within the CRTF that may themselves (or in varied combinations) predict positive student outcomes.

Theoretical Framework

Teacher effectiveness has become an important issue in today's current political landscape. While historically employee effectiveness has been an area of interest for many years with a wealth of research behind its claims, teacher effectiveness is a relatively new body of literature that is lacking empirical research studies on what characteristics determine good teaching and student improvement (Kyriakides, Demetriou, & Charalambous, 2006). While there is significant research linking teachers to their student's achievement (Darling-Hammond, 2000; MET Project, 2013; Wright et al., 1997), a gap in research exists between this knowledge and what actually constitutes an effective teacher. For the purposes of this study, teacher effectiveness will be defined as a teacher's characteristics that impact a student's achievement.

Teacher effectiveness can be measured in a multitude of ways and should therefore be based on multiple measurements. The work of the MET Project evaluated over 3,000 teachers and their organizations' methods of teacher evaluation to determine the impact of multiple-measures on student achievement gains and student perception. Their findings, over a 3-year period, clearly revealed that effective teaching can be measured with multiple measures and that the weighting of such measures makes a difference. In addition, multiple evaluators tend to increase inter-rater reliability (MET Project, 2013). Furthermore the work of Kyriakides et al. (2006) sought out to determine the most important characteristics that a teacher might perceive to be part of a teacher effectiveness program. The survey results showed that teachers consider most of the criteria presented to them from current teacher effectiveness research. In addition, clusters of characteristics emerged, determining new criteria for teacher effectiveness not initially determined by teacher effectiveness research.

The CRTF is Green Dot's multiple-measure evaluation rubric. It incorporates a teacher's planning and execution of instruction and the instructional environment, his/her ability to reflect and seek out growth opportunities, stakeholder feedback (student and parent), peer feedback, student growth (SGP), and school growth (school SGP).

Purpose of the Study

The purpose of this study was to determine whether or not Green Dot's CRTF could account, in whole or in part, for differences in student outcomes. In so doing, the study sought to determine the validity of Green Dot's teacher effectiveness program as predictor of student achievement outcomes.

Research Questions

This study was guided by the following research questions:

1. Are there differences in student outcomes (e.g., GPA and CST scores) based on a teacher's overall teacher effectiveness score?
2. Are there differences in student outcomes (e.g., GPA, CST, and SGP scores) based on a teacher's overall teacher observation score?
3. Are there differences in student outcomes (e.g., GPA, CST, and SGP scores) based on a teacher's score in a particular CRTF Domain?
4. Are there other observable constructs other than the five Domains identified by CRTF designers that can be arrived at through statistical analysis methods such as factor analysis?

5. Are there differences in student outcomes (e.g., GPA, CST, and SGP scores) based on a teacher's score in a particular CRTF construct arrived at through factor analysis?
6. Are there indicators (or combinations of indicators) that can predict positive student outcomes?

Delimitations and Limitations

Because this study was based on the effectiveness of the CRTF rubric and process at Green Dot, the following organizational procedures and policy limitations existed regarding the data set:

- Date range of data: The data included in the study came from the 2012-2013 school year.
- Attendance rules: Only students who were present 85% of the time from the first Wednesday in October 2012 to the end of the CST testing window were included in the data set. In addition, all teachers who missed more than 20 days of instruction were excluded from the data.
- Credentialed teachers only: Teachers who held valid teaching credentials (preliminary, CLEAR, or emergency) were included in the study. Long-term substitutes' data were excluded.
- Tested and non-tested teachers only: *Tested* teachers are teachers who taught courses culminating in their students participating in the California Standards Test (CST) at the end of the year. Tested teachers' students completed and received a score in one or more of the CSTs (e.g., English language arts, math, history, or science) and were eligible to receive an individual SGP measure. *Non-tested* teachers were also included

in the data set. These teachers taught courses (e.g., electives) that did not culminate in a standardized state test.

Definition of Terms

Academic Performance Index (API): A school's API score is a cross-sectional look at student achievement at the end of an academic year. The API is based on a number of measures: (a) results from the CST or modified version, and (b) California High School Exit Exam (CAHSEE) results. School academic growth is measured by calculating the change between a school's base API (results of the previous year) and the current year's API. API scores range from 200 to 1000 points.

California Standards Test (CST): The CSTs are standardized exams administered to students at the end of the academic year. The tests are designed to assess a student's proficiency in the California State Standards for English language arts, science, history, and math. CST results for each school are included in the school's API.

Student Growth Percentile (SGP): SGP is a statistical model used to determine a student's growth within a period of time (e.g., 1 academic school year). The SGP compares a student's achievement at the end of the year with comparable students who started at the same level at the beginning of the year. The members of this cohort of students are ranked from 1 to 100. A student's SGP in a particular subject will be that student's rank in the cohort. A teacher's SGP score can be determined from the median growth percentile (MGP), or the median score of all his/her students. One year of growth is typically noted by an SGP of 50. An SGP of 50 or less describes less than 1 year growth, while an SGP above 50 describes greater than 1 year's growth.

Teacher quality: Teacher quality is the all-encompassing term used broadly to describe the link between teacher activity and student learning. Teacher quality describes teacher qualifications, teacher characteristics, and teaching quality or practice (Goe, 2007).

Teacher qualifications and characteristics: Teacher qualifications include a teacher's educational background and experience including degrees, credentials, and certifications held. Teacher characteristics are broad and may include a teacher's background and credentials as well as other characteristics such as race, gender, self-efficacy, attitudes, religion, etc. (Goe, 2007).

Educator (teacher) effectiveness programs: Robust teacher evaluation programs that may rely on multiple measures to determine a teacher's effectiveness. A teacher's effectiveness is based on not only a teacher's ability to execute lessons in the classroom, but also his/her ability to lesson plan, incorporate student data informing practice, classroom management, student evaluations, peer evaluations, and more.

Teaching quality: A subset of teacher quality describing the practices a teacher engages in within the teaching practice. These include pedagogy, strategies used, classroom management, lesson planning, etc.

Teacher effectiveness: Teacher effectiveness describes the teacher's contributions to student learning (Goe, 2007). This may be measured through teacher effectiveness programs that may or may not include the use a student's performance on a standardized test.

Teacher Effectiveness Score (TES): The TES is a proprietary composite measure of teacher effectiveness developed by Green Dot Public Schools. The TES is based on a scale of 100-300. Teachers are assigned an effectiveness band based on their score. TES is calculated using weighted values from multiple measures: (a) classroom observation, (b) individual SGP,

(c) school wide SGP, (d) student survey, (e) Peer/360 survey, and (f) family survey. The teacher's subject area determines the weights of each component of the composite measure.

The College Ready Promise (TCRP): A coalition of several CMOs, TCRP is the foundation upon which Green Dot built the CRTF; it is the cornerstone of Green Dot's teacher and principal evaluation systems.

The College Ready Teaching Framework (CRTF): CRTF is a rubric based on Charlotte Danielson's (2011) *Framework for Teaching*. It defines the competencies of excellent teaching through five Domains: (a) data-driven planning and assessing student learning, (b) the classroom learning environment, (c) instruction, (d) developing professional practice, and (e) developing partnerships with family and community.

Value-Added Model (VAM): A form of teacher evaluation that seeks to determine a teacher's contribution to a student's achievement (the value-add of that teacher). VAMs typically measure the growth of students by comparing them to their scores from previous years. Statisticians predict the score of a student using the model and then compare this to the actual score. The difference between what was predicted and the actual score is said to be the value-add of the teacher and/or the school.

Chapter 2: Review of Literature

Background

NCLB. The No Child Left Behind Act of 2001 (NCLB) was the George W. Bush administration's reauthorization of the Elementary and Secondary Education Act (ESEA) enacted by Lyndon B. Johnson in 1965. Both ESEA and NCLB were concerned with school funding, student achievement, and most importantly, closing the achievement gap through ensuring equal access to education for all children. However, while NCLB continued the initial vision of ESEA, it also included an added focus on students' proficiency on state assessments and accountability measures that tied a school's status to its ability to meet AYP. The act required states to develop common state standards and create state assessments to measure student proficiency in these basic skills. The state would hold schools and districts accountable for educating their students and ensuring that they attained grade-level content standards. Students were required to take a yearly state assessment exam. Boldly, NCLB's core component and endgame is 100% proficiency in math and reading for all students by the end of the 2013-2014 school year (Paley, 2007).

School accountability. The reauthorization ties ESEA's flagship program, Title I, to increased accountability from the top down: state agencies, school districts, and school sites. As a result, in order to receive Title I funding, schools must meet accountability standards set by the states. Any school that fails to meet standards must undergo program improvement and remediation support provided by the local district. Schools that have undergone 3 or 4 years of remediation and have not shown improvement are further required to offer their students other choices of more effective schools. Funds were allocated, in these cases, primarily to provide other opportunities for parents to choose. This move's purpose was to ensure that the act meets its goals that no child loses an opportunity for quality education, thereby leaving no child behind (No Child Left Behind Act, 2002). The reauthorization essentially gave the federal government a larger role in public education, mandating the use of high-stakes exams, school accountability report cards, and teacher development tied to school funding.

Despite the strict attention to accountability and meeting standards, the act has been greatly criticized for its rigid, one-size-fits-all approach to education reform. A qualitative study by Mabry and Margolis (2006) examined NCLB's impact on educational practice. The study found that all administrators in the study population expressed anxiety regarding NCLB's unrealistic expectations. Mabry and Margolis concluded that while NCLB seemed to partially meet reforms in standards-based education, it did so to the detriment of its teachers in the form of increased attrition as well as teacher anxiety, and of its students, in the form of test anxiety and stress. Although NCLB policies promised reform through accountability, the goals were neither sustainable nor attainable.

Shift in focus. In 2010, the Obama administration released a *Blueprint for ESEA Reauthorization* (U.S. Department of Education, 2010), which outlined proposed changes to NCLB. These changes imparted much needed flexibility to NCLB, incorporating student growth measures in addition to school accountability and providing real rewards for high-poverty schools that showed growth and progress. The blueprint also allowed for the development of better assessments that paint a fuller picture of a student and school's growth and achievement. In addition, and most importantly related to this study, the changes sought to improve the profession of teaching through evaluation and development, providing teachers with expanded learning opportunities for professional development and collaboration among peers. The changes essentially removed the need for AYP instituted by the NCLB in favor of multiple-measure evaluation systems to develop highly qualified, highly effective teachers to contribute to school growth.

Though criticism was sharp because of the standards-based approach, NCLB ushered in an unprecedented era of accountability at the district and school levels. This accountability has turned its attention to the teacher, as evidence builds regarding the strong effects of teachers on student achievement. Today, there is an increased focus on a teacher's contribution to student achievement scores during end-of-the-year proficiency tests (Harris, 2011). Tracing back to the classic Coleman Report providing some of the earliest evidence linking teacher effects more closely to student achievement than any other available teaching resource (Coleman et al., 1966), research has been building behind the teacher as a determining factor of a student's ability to achieve (Darling-Hammond, 2000; Wright et al., 1997). The landscape of educational reform is again at a turning point. Many rubrics and frameworks have already been developed in an attempt to quantify the actions, behaviors, and thinking of an effective teacher, yet no consensus exists as to what specific characteristics truly determine the effectiveness of a teacher. Furthermore, if effective teachers produce effective and achieving students, a gap in research exists as to whether or not the teacher effectiveness systems and rubrics built to measure these effects can predict student outcomes accurately.

The purpose of this study was to investigate the validity of Green Dot's CRTF, a multiple measure evaluation tool that incorporates teacher evaluation, stakeholder feedback, and student growth, as an accurate measurement of student achievement measures. The subsequent literature review discusses research on teacher effectiveness, teacher evaluation, and student achievement measures.

Defining Teacher Effectiveness

There are many definitions of teacher effectiveness in the research literature (Goe, 2007; MET Project, 2013). For the purposes of this study, teacher effectiveness, as quantified by Green Dot's TES, can be defined as a teacher's impact on student achievement. This concurs with current research literature defining teacher effectiveness iterated subsequently. To date, however, there is currently no agreed upon definition of teacher effectiveness, let alone a standard measurement to determine this (Goe, 2007).

Goe's framework for teacher quality. Goe (2007) organized a framework for defining teacher quality that groups the many concepts and components of teacher effectiveness into three distinct categories: teacher inputs, teacher processes, and teacher outputs. Teacher inputs include what a teacher may bring to the teaching practice: his/her teacher qualifications as well as his/her teacher characteristics. Teacher qualifications describe a teacher's education, credentials, test scores and teaching experience. Teacher characteristics are defined as what a teacher believes, including his/her attitudes surrounding the teaching practice. Goe further defined a teacher's processes as the behaviors in which a teacher engages, such as the activities that occur inside his/her classroom, including interactions between the students and the teacher. Teaching practices include teaching strategies, lesson planning, classroom management, interactions with students, and a teacher's professional development choices. The combination of what a teacher brings to the teaching practice (inputs) and what a teacher does in practice (processes) seems to determine a teacher's effectiveness, defined by Goe as the *outcomes*, to some extent.

Inputs: Teacher certification, test scores, experience, and other qualifications. NCLB requires highly qualified teachers to be certificated, or certified by a state-approved credentialing agency, to teach a specific subject (No Child Left Behind Act, 2002). Most credentialing programs require teachers to have pedagogical and subject matter knowledge before they are credentialed to teach. Pre-service teachers complete coursework in teaching methods and pedagogy and are required to build experience through several hours of student teaching. This hypothesis assumes that teachers who are well prepared and possess appropriate certification may have a positive impact on student achievement. Current literature regarding teacher qualifications and certifications is mixed, however, providing little evidence strongly linking teacher qualifications to student achievement (Goe, 2007). Betts, Zau, and Rice (2003) attempted to correlate teacher qualifications (credentialing, level of education, subject matter knowledge) and student achievement. Their findings showed high variability among subject areas. Teachers with higher degrees such as a master's degree only contributed marginally towards student achievement, while teachers who possessed emergency credentials had much stronger correlations in mathematics than fully credentialed teachers with 10 or more years of teaching experience. This outcome runs seemingly counter to NCLB's assumptions of what constitutes a highly qualified teacher. In a study involving newly hired novice teachers in New York City, an improvement in TES of just one standard deviation was associated with an increase in student achievement comparable to the student achievement gains of veteran teachers, regardless of certification status (Kane, Rockoff, & Staiger, 2007). However, Goldhaber and Anthony (2007) presumed that unknown and uncontrolled-for confounding variables besides teacher qualifications may account for these effects.

In contrast, some link certification with student achievement. Darling-Hammond (2000) linked teacher qualifications to other student achievement; test scores such as the National Assessment for Educational Progress (NAEP) show promising correlations. In a study by Clotfelter, Ladd, and Vigdor (2006), National Board Certification was shown to have a statistically significant effect on reading scores. Teachers with alternate types of training, such as Teach for America graduates, had a similar impact on student achievement compared to their fully certified counterparts in mathematics. However, in other subjects they had less of an impact. It should be noted, however, that Teach for America graduates who went on to complete a full certification program were about as effective as teachers with traditional credentials (Darling-Hammond, Holtzman, Gatlin, & Heilig, 2005). Studies measuring teacher quality in terms of certification by the National Board for Professional Teaching Standards (NBPTS) consistently show that students of teachers who received board certification had larger achievement gains compared to students whose teachers had not pursued or had dropped out of the certification process (Cavalluzzo, 2004; Clotfelter, Ladd, & Vigdor, 2007; Goldhaber & Anthony, 2007). Furthermore, the effect of National Board certified teachers, those currently holding a certification as well as candidates, was strongest on students with low social economic status. This finding was also consistent with current National Board certified teachers who taught students receiving free or reduced lunch (Goldhaber & Anthony, 2007) compared to students who had not received such subsidies.

Teacher experience in the classroom is also thought to be linked to student achievement. There is substantial evidence of a direct positive correlation between a teacher's years of teaching experience and student achievement only up to the fifth year (Ferguson, 1991; Goe, 2007; Rockoff, 2004), with most teachers increasing in effectiveness during their first year of teaching (Kane et al., 2007). Monk's (1994) research on NAEP scores related to teachers' subject-matter expertise in four areas (coursework, major, degree, and experience) also found that student learning gains increase up to the fifth year, while teacher experience alone only contributed to student learning gains for students in the 11^{th} grade.

Input: Teacher characteristics. Teacher characteristics include the attitudes, beliefs, and immutable characteristics (e.g., race, ethnicity, gender, values) a teacher brings to the teaching practice (Goe, 2007). Goe (2007) asserted that research findings are largely mixed and more studies should be done to mitigate the contradictions between their findings. In one study, race matching, where a teacher's and student's race are identical, was shown to influence math and reading scores (Dee, 2004). However, in another study, a teacher's race, gender, and ethnicity lent no contribution to student achievement scores (Ehrenberg, Goldhaber, & Brewer, 1995). A study of teachers' beliefs involved in a school's internal social capital (e.g., trust and shared vision in the teaching practice) found that those beliefs influenced math and reading state proficiency tests (Leana & Pil, 2006).

Processes: Teacher practices. Teacher practice refers to the pedagogy, including strategies and behaviors, a teacher uses in the classroom to impact student learning. Tucker and Stronge (2005) asserted that "a reasonable consensus does exist on what effective teachers *do* to enhance student learning" (p. 242). It is widely known that most strategies for teaching are based on a wealth of learning theory research. Popular strategies such as *Think-Write-Pair-Share*, a collaborative learning strategy requiring students to work in groups to develop ideas, is based on a number of learning theories. The strategy is constructivist in that students *construct* ideas, forming new schemas and possibly the construction of new cognitive fields out of information gained in the environment (Derry, 1996). The strategy is also social in that, according to Vygotsky (1986), developing language through speech is an important process of a child's cognitive growth.

With decades of learning theory research behind the strategies teachers use in the classroom, a teacher's fidelity in delivering instruction through appropriate and polished pedagogy may affect student achievement directly. Studies on instructional practice models such as Madeline Hunter's seven-step instructional model not only resulted in teachers improving their instructional skills over a span of the study, but also various measures of student achievement showed promising results such as enhanced engagement time and increased gains in reading and math scores (Stallings, Robbins, Presbrey, & Scott, 1986). Furthermore, statistical correlations between teacher practices and student achievement in standardized testing assessments have shown promise. Holtzapple's (2003) work linking value-added measures of student growth to teacher evaluation ratings showed that students of teachers who earned lower evaluation ratings tended to have lower-than-predicted scores than students of teachers with the highest scores. Heneman, Milanowski, Kimball, and Odden (2006) examined the relationship between a common teacher evaluation system (the Danielson framework) and student achievement through value-added measures over a multi-year period at four school sites across the country. The study concluded that the specific instructional practices measured by the evaluation tool were related to higher test scores, suggesting that standards-based performance systems can have an impact on overall student achievement. Although the research is promising, the correlational effects, however, may be tempered by lack of alignment and pedagogical inconsistency between the teacher evaluation program and the state standards being assessed (Gallagher, 2004). Furthermore, a final determination of how these strategies should be implemented may still be subjective in nature (Friedman, 2006).

It has also been shown that student perception is impacted by the way teachers teach, which may indirectly affect student achievement gains. Wenglinsky (2000) found that students' perceived impact of teacher practices on their achievement outweighed the impact of professional development events for teachers. Student survey responses in Atlanta linked student achievement to effective instructional practice in mathematics and reading. Additionally, the study also showed a strong association between teacher expectations and student achievement (Frome, Lasater, & Cooney, 2005).

Teacher Evaluation

The combined effect of a teacher's inputs (qualifications, certifications, experience, and characteristics) and processes (practice) on a student's achievement is a teacher's *output* or effectiveness (Goe, 2007). To date, there is no current consensus as to what combinations of inputs and processes determine a teacher's *outcomes*. Teacher evaluation has had a complicated history and has evolved from a supervisory role to that of a coaching and reflective model. Two studies in particular outline current issues in teacher evaluation systems that sparked interest in the current study: the RAND study and *The Widget Effect* report.

The RAND study. The nature of evaluation in the early 1980s was criticized as being highly rigid, formulaic, and didactic. The RAND corporation studied 32 school districts to determine what systems of evaluation were being implemented across the country. According to the findings, the reflective nature of evaluations lacked enough specificity to support professional development in specific teaching pedagogy. Teachers felt that administrators were not competent enough to evaluate their teaching accurately. As a result, teachers were unresponsive to feedback and input from administrators. The study also uncovered possible reasons for this phenomenon. There was a lack of uniform evaluation systems across the districts that determined teacher competencies. Furthermore, administrators were not adequately trained or calibrated to systems of evaluation, causing scores to be inconsistent from one administrator to another (Wise, Darling-Hammond, & Bernstein, 1985). The study provided five key conclusions and recommendations for teacher evaluation systems moving forward. Teacher evaluation systems should:

1. Match the goals, management styles, and conception of teaching of the school district. Before implementing a teacher evaluation system, districts should take inventory of their mission and vision for education for alignment with the evaluation system. In addition, evaluation systems should not be highly prescriptive (Wise et al., 1985).
2. Be committed to and provide support for the evaluation system: (a) the evaluation system should be evaluated periodically for quality and evaluator competence, (b) adequate training should be provided for evaluators and teachers, and (c) resources pivotal to the evaluation system should be leveraged to support the system. For example, assistant principals may share the load of teacher evaluation, should the demands become too great (Wise et al., 1985).
3. A teacher evaluation system should match a school district's purpose for teacher evaluation (Wise et al., 1985).
4. Show the utility of a teacher evaluation system in order to generate commitment and adoption of the system. School districts must be transparent about the utility of their teacher effectiveness program in order to generate buy-in from all stakeholders. The resources used in the program (e.g., rubrics, resource repositories, online websites, etc.) should be linked directly to ensuring that the process is efficient, cost-effective, and valid (Wise et al., 1985).

5. Teacher involvement in the development of the teacher evaluation system should ensure that the system is legitimate and fair, as well as effective. Furthermore, teachers should be held accountable to instructional practices (Wise et al., 1985).

The Widget Effect. In 2009, The New Teacher Project's groundbreaking study of 15,000 teachers and 13,000 administrators in 12 school districts across four states found that school districts have failed to acknowledge, report, and/or act on differences in teacher performance. This observed act of indifference, where a teacher's performance had no bearing on decision-making for the sake of students, formed the basis of their study known as *The Widget Effect.* The report accused current teacher evaluation as "disrespectful to teachers," "indifferent to instructional ineffectiveness," and "gambling with the lives of students" (Weisberg et al., 2009, p. 4). Essentially, teachers are like *widgets*. Widgets are things that are considered typical and representative of something, such as a rear-view mirror is typical of a car. One would expect a rear view mirror to be the same from car to car. The study points out that teachers, like widgets, are assumed to be the same from classroom to classroom. Evidence of the effect can be seen in tenured teachers previously identified as ineffective continuing to teach without consequence or remediation. In contrast, while ineffective teachers continue to remain ineffective under the protections of tenure, excellent teachers are not recognized, rewarded, or developed further.

The Widget Effect argues that teacher evaluation is ineffective in that it does not provide adequate differentiation for teachers' evaluation and therefore evaluations tend to favor only satisfactory ratings. Evaluations are highly subjective and prescriptive; some even rely on checklist style tallies to determine the rating. In evaluation systems with binary options for ratings such as *satisfactory* and *unsatisfactory*, 99% of teachers were rated as satisfactory. When evaluators were provided multiple-ratings options, 94% of teachers were in the top two ratings scores and less than one percent of teachers were rated as unsatisfactory (Weisberg et al., 2009).

With a lack of differentiation in a teacher's evaluation, ratings are not valuable as a professional development tool. Current evaluation systems are not differentiated among teachers and therefore ignore teacher performance. When teacher performance is ignored, great teachers cannot be rewarded, ineffective teachers remain unchallenged, and all teachers cannot receive specific and targeted support. The study further uncovered the perceptions of teachers regarding their evaluation: 73% of teachers reported that their most recent evaluation did not identify areas for development, with only 45% of the teachers reporting that they did receive adequate support to improve (Weisberg et al., 2009).

The study concluded that there was a need for a fresh envisioning of teacher evaluation that differentiates and takes into consideration the full range of what teachers do. The report provided four recommendations to reverse the Widget Effect:

1. Adopt a teacher evaluation system that is comprehensive, differentiated, and fair. To do this, the report suggested clear performance standards, multiple ratings options (e.g., performance band indicators similar to Danielson's Framework), frequent feedback to teachers, and monitoring of administrator judgments on evaluations.
2. Proper training in use of the evaluation tool. Administrators and evaluators must be trained properly in order to remain fair and consistent in their evaluations and to achieve the objectives of the evaluation system described in item 1.
3. The teacher evaluation system should be integrated with human capital policies and functions. Data provided through the evaluation should inform hiring procedures,

teaching assignment, professional development and intervention, compensation, and dismissal procedures.

4. As a result of an effective system of evaluation, dismissal policies should be adopted that provide lower-stakes options for exiting teachers. This process must also be transparent, fair, and equitable, and must involve due process (Weisberg et al., 2009).

Multiple measure evaluation. Teacher evaluation is arguably as complex as defending a legal case or performing a surgical procedure (American Federation of Teachers, 2013) and as a result should be measured with multiple methods ("Multiple Choices," 2013). Therefore, according to Goe, Biggers, and Croft, (2012) multiple-measure evaluations can paint a more complete picture of a teacher's effectiveness across multiple Domains of measurement beyond only student achievement data. In 1929, William Wetzel (as cited in Marzano, Frontier, & Livingston, 2011) rejected Cubberley's (1916) factory metaphor of education in favor of a different model of evaluation where classroom observation of practice was as important as student achievement scores. He posited three factors of importance in evaluation: student achievement through reliable measures, the construction of clear and measureable course objectives, and determining student ability through aptitude testing (Marzano et al., 2011). The MET Project (2013) set out to address some of the concerns presented by the Widget Effect by developing and testing comprehensive, multiple-measure teacher effectiveness systems. By improving the quality of information about teacher effectiveness systems to all stakeholders, the MET Project's goal and mission is to equip schools and districts with better tools to design fair and reliable teacher evaluation systems that can meet the needs of teachers and ultimately the success of students. The MET Project found that multiple measure evaluations resulted in more consistent ratings than student achievement alone. In addition, they found that the results are more stable from year to year when classroom observations, student feedback, and student achievement gains are included in the measure compared to student achievement alone.

According to RAND (as cited in "Multiple Choices," 2013), the following measures are valid options for measuring teacher effectiveness:

1. Student achievement test scores are an efficient way of measuring the value-add of a teacher and/or student growth.
2. Classroom observation can measure teacher practices and pedagogy through direct observation, lesson plan analysis, and artifact analysis.
3. Student surveys provide feedback from all stakeholders including but not limited to student engagement, teacher-student relationships, and perceived teaching ability from students' perspective.

The composite teacher effectiveness score (TES) suggests predictability. The MET Study demonstrated that a composite TES can predict a teacher's success better than some teacher characteristics (e.g., teacher's years of experience and education level; MET Project, 2012). A study by Dartmouth College and RAND Corporation used MET Study data to compare the implications of differently weighted components of the composite score (MET Project, 2012). The research sought to determine whether or not different weightings of the components of TES were associated with predicting better student gains as well as the trade-offs with various models. Four models were suggested: (a) control model for maximum accuracy in predicting gains on state tests: 81% student test gains with 17% student survey, and two percent observation; (b) 50% student test gains with 25% survey and observation score; (c) 50% on observations with 25% for student test gains and surveys; and (d) equal weights for all. The results showed that models b and c, where state test gains weighting were at 50% and 30%, respectively, still had similar predictive power compared to model a. There were tradeoffs, however; while model a exhibited the most predictive power, it was the least reliable of all composites and had the worst predictability in higher-order testing. Furthermore, models that placed heavier weightings on classroom observations had the least correlations with state test gains and higher-order tests. The tradeoff for these low correlations was that these models were most reliable among the tested models.

History of teacher evaluation. Teacher evaluation has had a rich history, beginning in the 1700s with the teacher's role as a community servant. The clergy took up the role as supervisors, having the power to determine what criteria led to effective instruction, simply because of their level of education and presumed ability to teach religious studies. The realization that supervisors required subject matter knowledge and teaching skills in addition to teaching pedagogy in order to improve instruction and teacher expertise (Tracy, 1995) caused a major shift from subjective evaluations to more precise and measurable systems of evaluations. This shift in thinking eventually led the movement to view education as a science of management. Borrowing from the work of Frederick Taylor (1911), the work of a teacher could essentially be broken down into discrete tasks that could be evaluated and measured for best practices offering the greatest return on investment. Furthermore, Cubberley (1916) viewed schools as factories of children's minds that were shaped by the factory workers: the teachers. Therefore, like a factory, scientific and measurable systems of quality needed to be in place to properly observe and supervise teachers to ensure school success and quality. The following section discusses the major components of teacher evaluation.

Goldhammer's five phases of supervision. In the 1950s, a systematic approach to teacher supervision involving a clinical cycle of supervision became popular. Goldhammer (1969) proposed five phases of supervision: (a) pre-observation conference, (b) observation, (c) analysis, (d) supervision conference, and (e) analysis of the analysis. The cycle was originally intended to develop a purposeful relationship between the teacher and the supervisor or administrator, much like the learning that occurs in teaching hospitals where resident physicians are under the tutelage of a master physician. Goldhammer argued that such a protocol was designed as a method of evaluating, through a holistic lens, the practice of teaching: the interactions between teacher and his/her students (Marzano et al., 2011). The process of dialogue between an observer and a teacher was intended to be instructive and coaching in nature, disclosing effective instructional practices throughout the coaching process. Indeed, in the mid-1980s, the focus moved from teacher evaluations as being strictly rigid, prescriptive, and scientific, to a focus on evaluation for the purpose of teacher professional development. Administrators would not simply serve as evaluators observing and completing checklists; rather, they would be focused on developing effective teachers and shifting from supervisory roles to engaging in coaching and evaluation. As such, the process became more reflective in nature as a means towards the end goal of student improvement through improved instruction. Furthermore, Carl Glickman (1985) offered an approach to supporting teachers through innovation in professional development (group and individual), direct assistance to teachers, goal setting and attainment, and enhanced curriculum development.

Moving to more effective lesson planning. A major innovation in teacher evaluation and support came in the form of the evaluation of the lesson plan or cycle. In 1984, Madeline Hunter developed an instructional model identifying key components of lesson design. The seven-step instructional model known as *The Hunter Method* or the *Madeline Hunter Direct Instruction Model* of mastery teaching quickly became the language of effective instruction against which effective lesson planning and execution were evaluated (Marzano et al., 2011). The seven steps in an effective lesson plan include:

1. The anticipatory set: A warm-up activity that focuses students on the learning that will occur in the subsequent lesson. The anticipatory set can also be used to diagnose misconceptions or to activate prior knowledge.
2. The objective and purpose: Following Wetzel's (as cited in Marzano, Frontier, & Livingston, 2011) three-factor model of evaluation in 1929, clearly-written objectives outlining the expectations and purpose for the subsequent lesson help both students and teacher stay accountable to the goals of the lesson.
3. Input: The process whereby students acquire new information. This can occur via presentation or direct instruction.
4. Modeling: Students see an example and/or exemplars of the product based on the objectives of the course and the input(s) provided.
5. Checking for understanding: A teacher makes effective instructional decisions throughout the lesson cycle through periodic formative assessments known as *checks for understanding*. Feedback from checks for understanding informs the subsequent instruction.
6. Guided practice: Students practice the newly learned knowledge or skill with supervision from the teacher.

7. Independent practice: Once a skill or knowledge has had adequate practice time, students can complete the task without supervision from the teacher. This proving behavior of the lesson offers evidence that the objective has been met (Hunter, 1982).

Classroom observation. In 1996, Charlotte Danielson embarked on a seminal work known as *Enhancing Professional Practice: A Framework for Teaching*. The work offered a comprehensive rubric for evaluating the dynamic teaching practices of a classroom teacher (Danielson, 2007). The framework was based on Danielson's work with the Educational Testing Service evaluating pre-service teachers for competence in teaching. It attempted to capture the teaching practice in its full complexity by considering specific teaching practice indicators across four major Domains: (a) planning and preparation (Domain 1), (b) the classroom environment (Domain 2), (c) instruction (Domain 3), and (d) professional responsibilities (Domain 4). Each Domain is broken down into distinct competency indicators that attempt to describe the practices, dispositions, knowledge, and skills necessary for effective teaching. A teacher's effectiveness can be determined by rating him/her against four levels of performance (unsatisfactory, basic, proficient, and distinguished) on each performance indicator (Danielson, 2007).

An early study involving the framework in a low-stakes, non-evaluative teacher evaluation system showed promising correlations between a teacher's performance on the evaluation and student achievement; however, results were not consistent across all cases. This was perhaps due to the low-stakes nature of the evaluation or the non-specificity and/or sensitivity of the instrument when evaluating teachers on pedagogy and content (Kimball, White, Milanowski, & Borman, 2004). Furthermore, numerous studies examining full-fledged teacher evaluation systems across the country have validated the tool as a valid predictor of student achievement (Heneman et al., 2006; Holtzapple, 2003; Kimball et al., 2004). As a result, today the Danielson model is widely regarded as the reference point for most teacher evaluation rubrics, including the CRTF, the rubric in the current study.

Student achievement. One of the most direct effects of teacher effectiveness can be seen in student achievement. A number of student achievement metrics can be used to demonstrate student proficiency. The current study used GPA, standardized testing scores (in the form of CSTs), and SGP. The following section outlines the pros and cons of each measure selected in the study as indicators of student achievement and as components of a teacher evaluation system.

GPA. GPA has been used as a gauge of student proficiency in high schools for quite some time. In general, high GPA denotes achievement across all courses a student takes in a given marking period. GPA has been shown to predict student success in college weakly. In one study examining the ACT Composite test and high school GPA as predictors for college success, GPA was found to be more valuable than college admission test scores alone when admission engaged in low selectivity (Sawyer, 2013). Furthermore, Geiser and Studley (2002) found that GPA in college-preparatory classes was the best predictor (by itself, accounted for 17% of the variation) of freshman grades in a sample of 80,000 students admitted to the University of California. Key findings in a follow-up study showed that high school GPA continued to be a strong 4-year predictor of student grades throughout college for most major academic fields. Additionally, GPA had less adverse impact on minority and underrepresented student groups than the SAT (Geisler & Santelices, 2007). Furthermore, GPA, when coupled with the results of the ACT composite test, is an accurate predictor of student GPA as a first year student in college (Noble & Sawyer, 2002; Sanchez, 2013).

Alternatively, criticisms of GPA as a reliable measurement for students (and indirectly for teachers) specifically cite concerns about grade inflation and the absence of grading standards across all schools. For example, as Camara, Kimmel, Scheuneman, and Sawtell (2003) noted that grades have been inflating steadily over the course of a 25-year period. According to the Cooperative Institutional Research Program (as cited in Camara et al., 2003), a staggering 42.9% of college freshmen in the year 2000 reported GPAs of 4.0 or above. One of the possible factors causing grade inflation may be that grading standards have changed, edging away from a standard, normal distribution to a less stringent, more subjective form of grading rewarding effort rather than standards.

Standardized testing. The purpose of state assessments, according to the U.S. Department of Education (2004), is to "provide an independent insight into each child's progress, as well as each school's" (para. 1). As a metric for student achievement, standardized testing provides a reliable way to determine proficiency that is not subject to the inflation or inconsistency seen with GPA. According to a 2011 research summary analyzing 100 years of teaching research, 93% of all studies on the use of student testing have found a positive effect of standardized testing on student achievement. Furthermore, standardized testing is a reliable and objective method of student achievement in that it ensures that results are not skewed or biased by separate organizations and districts (Phelps, 2011).

Conversely, many issues arise regarding the use of student achievement tests. Achievement tests are limited in that they only measure a small portion of what makes education meaningful (Strauss, 2011) and they are administered infrequently. While standardized testing may be valid for representing student proficiency at the end of the year specifically, used alone, it may not be a valid measure of teacher effectiveness (Betebenner, 2011). Goe, Bell, and Little

(2008) suggest serious implications exist in the literature regarding the use of student achievement data as the sole determinant of teacher effectiveness. Using student achievement data alone assumes that all students are the same and that factors directly attributable to the teacher, such as teaching characteristics and pedagogy, as well as factors attributable to the school environment, have no bearing on a student's development and therefore no bearing on his/her performance. Furthermore, if a teacher's effectiveness is to be measured using student achievement data, the student assessment must be valid and reliable, useful for diagnosis, and allow for equitable access to the assessment by all learners. Currently, no assessments meeting those requirements exist.

Arguing a "clear and undeniable link that exists between teacher effectiveness and student learning," Tucker and Stronge (2005, p. 102) championed the use of student achievement data as a source of feedback for the teaching practice, as well as the effectiveness of teachers, administrators, and the school as a whole. Tucker and Stronge asserted that student achievement data may have four important implications in determining teacher effectiveness:

1. Standardized testing is an objective measure of teacher effectiveness. Unlike teacher observation, which can introduce observer bias into the evaluation process, standardized testing is highly objective.
2. Standardized testing can provide meaningful feedback to teachers being evaluated. Because standardized testing is highly objective, data can speak for itself, allowing teachers to "look honestly at their weakness and strengths" (Howard & McColskey, 2001, p. 49).
3. Standardized testing data can serve as a barometer of success. Feedback in terms of student achievement data can be used as intrinsic motivation to better one's practice.

4. Standardized testing is already an integral facet of instruction.

Value add and student achievement. Student achievement data alone as a measure of teacher effectiveness has limitations. While it may be a fairly adequate measure of student proficiency, standardized tests cannot account for the effectiveness of teachers by itself since student achievement scores alone may be affected by factors such as socio-economic status, ethnicity, and language bias. To counter this, VAMs of student achievement data have been adopted in an attempt to measure a teacher's contribution to student learning without the confounding effects of non-educational factors. Current policy and literature define a teacher's effectiveness as his/her impact on student growth. With these factors removed from the mix, one can assume that the value-added effect of a teacher on a student, class, or school can be measured.

VAMs employ linear projections as statistical analyses of longitudinal data. Statisticians use a student's prior test scores to predict the student's future test scores. This assumes that a student, on average, performs just as well every year as the student has in the previous years. The student's predicted score is then compared with the student's actual score. The difference between a student's actual and predicted score is presumed to be the result of the contribution of the teacher and/or school to a student's achievement. Essentially, the difference reflects the value-add of the teacher and/or school.

VAMs, according to McCaffrey et al. (2003), have also been show to expose very large differences in effectiveness among teachers. Such differences can be further studied and linked to teacher characteristics and practices informing professional development in ways unrealized until now. The use of current VAMs for determining teacher effectiveness, however, is still controversial. Sanders and Horn (1998) contended that variables such as poverty and cultural school climate are inherent in a student's prior test scores used to predict future achievement. Other variables such as class compositions and other school effects that are not controlled for may also confound students' prior scores (McCaffrey et al., 2003).

Growth measures and SGP. In addition to value-add, growth modeling has attempted to compare students to other students within similar cohorts. As of 2008, Slaughter (2008) noted that 11 states have adopted growth modeling into their AYP tracking plans. The Colorado Department of Education has adopted one such statistical model known as the SGP. SGP employs a normative approach to evaluating growth in student achievement. Unlike VAMs that evaluate a student's growth-to-standard based on a student's own longitudinal record of achievement (Betebenner, 2009), SGP is based on normative criterion-referenced growth. In other words, SGP compares how much a student has grown compared to similar students who started the year at the same level. The model considers normative growth instead of absolute growth. SGP has also been shown to control for the residual non-educational factors for which VAMs have been criticized (e.g., socio-economic status, race, and ethnicity).

It is also important to note, however, that SGP does not arrive at the table without criticism, as Briggs and Betebenner (2009) noted that SGP is a descriptive measure, evaluating growth at the student level in aggregate. According to Baker (2011), SGP was never intended to estimate the school or teacher's effect on the student's growth.

Perception survey. The practice of using perceptual data is being incorporated into teacher evaluation systems as a means of providing more meaningful feedback to teacher evaluation systems. Student perception surveys provide feedback from the individuals who are most directly affected by teacher effects. The MET Project (2012) asserted that student survey data produce more consistent results than classroom observation alone due simply to the survey respondent's more intimate experience of his/her teacher's ability to teach. According to the MET Project's findings, student perception results with the Tripod survey showed positive relationships between a teacher's Tripod percentile rank and his/her VAMs in math and English language arts.

Furthermore, as with any other evaluation instrument, stakeholder surveys must provide valid and reliable data. The MET Project findings assert four requirements to student perception surveys: (a) they should measure teacher practices and the learning environment that teachers create, (b) they should ensure accuracy, requiring students to provide honest responses, (c) they should be reliable, and (d) they should support improvement by providing valuable and actionable feedback to the teacher receiving them (MET Project, 2012).

Summary

Teacher evaluation has had a long history of developing measures of effective teaching through considering multiple measures such as classroom observation, lesson plan evaluation, stakeholder feedback, and student achievement scores. Many of these measures by themselves show some relationship with student achievement and teacher effectiveness. Furthermore, the MET Project, the Widget Effect*,* and the RAND report have all provided recommendations as to what measures should be included in an effective teacher evaluation system. Green Dot's teacher evaluation system was developed after considering the recommendations of the MET Project and involves multiple measures in determining a teacher's composite TES. While many of the recommendations were followed, no formal evaluation of the program has yet been conducted. This quantitative study aimed to validate Green Dot's teacher evaluation system through determining whether or not a relationship exists between teachers' scores on evaluation system measures (overall, Domain scores, observation scores, and indicator scores) and student outcomes. The implications of this research may inform future professional development on indicators or Domains that are most related to student achievement. Furthermore, Green Dot may also consider rethinking or evaluating the rubric for indicators or Domains that are not related to student achievement.

Chapter 3: Methodology

This chapter presents the methodology used in the current investigation. The study examined the validity of the CRTF as a valid predictor of student achievement outcomes. Of particular interest is whether or not the CRTF affects measurable student outcomes in SGP, CST scores, and student GPA. The study sought to determine whether or not correlations exist between a teacher's overall TES on the CRTF and student outcomes. Current research on similar rubrics suggests no strong positive correlations between teacher effectiveness measures and student outcomes. The significance of this study was the hope that should relationships be found to exist between indicators, pre-existing Domains, or emerging factors, the CRTF would become a more effective tool with implications for professional development and hiring practices leading to better student achievement.

Research Design

This study employed a quantitative research design focusing on a secondary analysis of existing data. The study employed correlational analysis to determine whether or not there is a relationship between overall TES and student outcomes (e.g., SGP for growth, CST scores, and GPA) as well as teacher observation score (TObs) and student outcomes as measured by Green Dot's teacher effectiveness system. Correlational analysis was chosen because of its ability to show how strongly pairs of variables (TES and TObs versus student outcomes) are related by the strength of the Pearson product-moment coefficient (*r*; Gertsman, n.d.; Taylor, 1990). Pearson coefficient values of less than or equal to 0.3 are generally considered to be weak correlation, while values of 0.3-0.7 are moderate and 0.7-1.0 represent strong correlations (Gertsman, n.d.). R-squared values, known as the *coefficient of determination,* represent the percent of the variation in the observed values (student outcomes) that can be accounted for in the variations of the independent values (TES or TObs). R-squared, then, can provide insight into the significance of the correlation (R. Taylor, 1990).

Furthermore, the study employed factor analysis to determine if there were other constructs or *factors* other than the theoretical CRTF Domains that were created by the Charlotte Danielson and Green Dot. Factor analysis was deemed a valid statistical method for this purpose as it attempts to identify additional constructs or Domains that emerge from linear correlations between the observed variables (in the present study, the CRTF indicators) and a smaller number of unobserved variables (also known as Factors). Correlation and R-squared values were then used to determine if these new constructs were associated with and/or could account for the variation in the observed student achievement data (Tryfos, 1998).

Multiple regression analysis was also used for predictive models with multiple independent variables. Multiple regression analysis provides a working model based on the slope of a line of best fit through the plotted variables. This process was used in this case to determine if certain indicators could predict student outcomes by themselves or in combination. An R-squared value was calculated in this case (as the *coefficient of multiple determination*) to determine how closely the data fit the theoretical model by reporting the percentage of variation attributed to the model.

Description and Selection of the Subject Sample

The population considered for this study consisted of teachers and students in Green Dot during the 2012-2013 academic year. Teacher and student data were collected from four middle school campuses and 16 high school campuses operated by Green Dot in the greater Los Angeles area. The teacher data in the study were part of Green Dot's educator effectiveness program and were collected under requirement by collective bargaining agreement. Student data were collected as a part of a student's academic transcript. In addition, the data were required by the state in terms of NCLB data metrics through state-mandated testing.

Teachers. The teacher participants in this study included 239 ninth through 11th grade high school teachers and 51 sixth through eighth grade teachers from Green Dot's 18 schools. Most teachers served, on average, 28 students per class period. Eighty-eight percent of teachers had teaching credentials in various subject disciplines. Teachers who did not meet the attendance requirement, missing more than 20 days of instruction, were excluded from the data set. Both *tested* and *non-tested* teachers at Green Dot were used in the sample (see *Teacher Effectiveness Score* section). Data from teacher effectiveness were extracted from the 2012-2013 Teacher Effectiveness Score Cards (TESC). The TESC represents the culmination of Green Dot's teacher effectiveness program providing the teacher with all the data points comprising a his/her TES: stakeholder feedback, CRTF average score, SGP (individual and/or school), and compliance scores (for SpEd teachers).

Students. Student data from 6,603 students from grades 6-11 were collected through Green Dot's student information system, PowerSchool. GPA was recorded as an average of student grades twice per year (once per semester) and placed on a student's permanent academic transcript. A student's CST scores were assessed on or after the 85th (+/- 20) day of instruction as required by the state to comply with NCLB and state and federal accountability policies. Finally, Green Dot calculates a student's SGP for a specific subject on a yearly basis. All students were enrolled full time at one of Green Dot's middle or high schools. Students who did not meet the student attendance rule of 85% attendance were excluded from the population.

Ninety-two percent of all students qualified for free and reduced lunch status. Seventy-eight percent of students were of Latino or Hispanic descent, 20% were of African-American descent, and the remaining two percent belonged to other ethnic groups.

Data Collection

This study was based on a secondary analysis of data obtained from two major sources: Green Dot Public Schools Educator Effectiveness program and Green Dot Public Schools student information system (SIS). All data were previously collected as per organization and collective bargaining agreements during the 2012-2013 school year.

Student Information System and GPA. Student grade data was collected using Green Dot's central SIS, PowerSchool. The platform allows teachers to take daily attendance, record disciplinary log entries, access historical grades and GPA, and, most importantly, record the grades for each assignment, which ultimately leads to the calculation of final grades. Final grades are then recorded at the end of each semester cycle (20 weeks). The calculation of final grades in a class depends on how a teacher sets up his/her grade book. By default, grading is not weighted and is calculated simply by dividing the total points a student has earned throughout the marking period by the total possible points available, yielding a percentage score. However, some teachers opt to weigh grades differently according to category weights and calculate final grades based on a percentage of each category weights. Once a final grade is finalized at the end of a semester, GPA is calculated by converting final letter grades into numerical grade points using a 4-point scale. An A converts to a 4 and a D converts to a 1. Letter grades of F receive no grade points. The final semester GPA was calculated by taking the average of the grade points.

California Standards Test scores. CST scores were collected for four exams: English Language Arts (ELA), science, history, and math. Students took the corresponding CST that mapped to the current core course they took during the school year. The exam was administered to all students in the sample within a 2-4-week testing window in the spring of 2013. Students provided responses to test questions via a bubble-type, computer scanned answer document under strict and stringent testing conditions required by Green Dot. Teachers only proctor test subjects they do not immediately teach. In other words, teachers who teach math cannot proctor a math CST. To ensure fidelity of the testing environment including reliability of scores and test security, all teachers are also required to sign affidavits acknowledging their compliance with testing and security procedures. Completed exams are returned to the testing vendor for grading immediately via FedEx. Responses are then scanned by the vendor and the results returned to Green Dot by August 2013.

CRTF observation scores. Data from 29 discrete indicators labeled 1.1.A through 4.1.B were collected as part of the formal observation process cycle twice a year. This cycle consisted of a lesson plan review, classroom observation, teacher reflection, and student artifacts. Administrators gathered evidence, tagged the related to corresponding indicators in the CRTF rubric, and ultimately provided a final rating against the CRTF rubric. A web application known as Bloomboard facilitated this process for both teachers and administrators. Once finalized, a teacher's CRTF score was used in the calculation of the TES.

Student feedback. According to Green Dot's 2012 through 2013 collective bargaining contract, students are required to evaluate their teachers twice a year through a survey. The 2012-2013 survey captured a student's perception of a teacher's proficiency in Domains 2 and 3 of the CRTF. The survey was administered from November 5-16, 2013 and a second time from April 22 through May 3, 2014, during a 2-week testing window. Responses were recorded using a bubble-type answer document that was scanned into the Scantron Survey Tracker database by

Green Dot's Knowledge Management (KM) department. Students completed a maximum of two surveys on two of their teachers. The evaluated teachers were randomly assigned to their student evaluators by a computer algorithm. Each teacher was evaluated one to 65 times, depending on the number of students a teacher had. All survey scores were averaged to calculate a teacher's overall student survey score for each semester. A teacher's overall student survey score for the year is a weighted average comprising 40% of semester one and 60% of semester two survey scores.

360 feedback. Teachers were required to rate three randomly selected peers from their school on Domains 4 and 5 of the CRTF. In addition, teachers also completed the same survey on themselves (though self-ratings were not used in the calculation of the TES and are therefore irrelevant to the study). Responses were also recorded using a bubble-type answer document that was scanned into the Scantron Survey Tracker database by Green Dot's KM department.

Parent survey. Parent/family surveys were administered once in 2012-2013 as an assessment of Domain 5. Responses were also recorded using a bubble-type answer document that was scanned into the Scantron Survey Tracker database by Green Dot's KM department. An overall average was provided for each school. The calculation of the TES involves this average score for each teacher, regardless of whether or not the teacher was a tested, non-tested, or SpEd teacher.

SGP. Using the CST scores (e.g., scale scores and test type) from 2011-2012 and 2012-2013, a third party statistical firm calculated student SGP scores per subject area. The method employed involves the use of *R*, a statistical platform with the SGP package extension (Betebenner, 2009). Data from 2011-2012 served as the beginning or prior score, establishing the students' peer group and pathway. Population data included data from Green Dot as well as the larger Los Angeles Unified School District (LAUSD). CST data from 2012-2013 (e.g., scale scores and test type) formed the comparison or ending score for the SGP calculation. Once scores were calculated they were returned to Green Dot in November 2013.

Data warehouse. Green Dot's data warehouse is a relational database stored on Microsoft SQL Server. The data warehouse stages all student, teacher, and employee data. Data are linked to students and teachers via student number and employee ID, respectively. Student data from PowerSchool are synchronized with Green Dot's data warehouse on a nightly basis. CST scores, once scored by the testing vendor, are cleaned and processed by Green Dot's KM department before being uploaded into the data warehouse. Teacher CRTF observation data housed in Bloomboard are synchronized with the data warehouse on a nightly basis. Survey scores are uploaded from Scantron Survey Tracker once they are cleaned and processed by KM.

Data delivery and confidentiality. Stringent protocols were followed to ensure the confidentiality and anonymity of the data provided. Anonymous student and teacher data were provided to the researcher in the form of a Microsoft Excel file from Green Dot's database administrator (DBA). Students were identified in the file as a nine-digit, randomly generated ID code ending in the letter S. Teachers were also identified as a nine-digit, randomly generated ID code ending in the letter T. The junction table containing the key to link the student number and employee IDs of the subjects to their identities was kept in a secure human resources table within the data warehouse and was not accessible to non-HR personnel. At the conclusion of the study, the key table was deleted, rendering it impossible for the researcher to identify student or teacher participants.

Instrumentation

The College Ready Teaching Framework. The CRTF is a rubric defining the competencies of excellent teaching. The rubric is made up of 5 Domains: (a) data-driven planning and assessing student learning, (b) the classroom learning environment, (c) instruction, (d) developing professional practice, and (e) developing partnerships with family and community. Domains are broken up into sub-Domains or standards. Various skill-level competencies or indicators make up each standard. Each indicator is assigned a rubric score denoting levels of proficiency. A score of a 1 indicates that the teacher does not meet the standard. A level 2 represents that the indicator only partially meets the standard. A level 3 denotes a teacher meeting the standard, whereas a level 4 means that the behavior observed exemplifies the standard (see Figure 2). It should be noted that the five Domains measured on the instrument were identified by the designers of the instrument. No analyses (such as factor or principal component analysis) have been conducted to verify these Domains.

Teacher Effectiveness Score (TES). Teacher effectiveness and performance were measured using Green Dot's TES, a composite measure of teacher observation, SGP, and stakeholder feedback. Teachers at Green Dot fall into one of three types: a teacher who teaches a course that is directly mapped to a CST exam (e.g., math, ELA, history, or science) and who has 100 or more valid SGP scores over a 2-year period falls into the Tested Teacher type. Those who teach courses that do not map to a CST exam (e.g., some elective courses, Art, Music, etc.) or do not have 100 or more valid SGP scores over a 2-year period are considered Non-Tested teachers. Finally teachers who teach special education students exclusively (e.g., students who have an Individualized Education Plan and require accommodations and modifications) are considered part of the SpEd teacher type. SpEd teachers were not included in the study.

Domain 1: *Data-Driven Planning and Assessing Student Learning*

Standard	Indicators	Does Not Meet Standard Level I	Partially Meets Standard Level II	Meets Standard Level III	Exemplifies Standard Level IV	Indicator At-A-Glance
1.1 Establish standards-based learning objectives and assessments	A) Selection of learning objectives	Learning objective(s) are missing a specific level of cognition or content. AND Learning objective(s) are misaligned (do not progress toward mastery of content standards).	Learning objective(s) are missing either a specific level of cognition or content. OR Learning objective(s) are misaligned (do not progress toward mastery of content standards).	Learning objective(s) include both specific levels of cognition and content. AND Learning objective(s) are aligned to and progress toward mastery of content standards.	*All of level 3 and...* Learning objective(s) exceed level of cognition or increase level of challenge required by content standards.	• *Do the objective(s) contain level of cognition and content?* • *Are the objective(s) at an appropriate level of rigor and scaffolds toward success on standard(s)?* *NOTE: Proving behavior is no longer assessed in this indicator. Now, it is assessed in 1.1B.*
	B) Measurability of learning objectives	Proving behavior does not measure the learning objective(s).	Proving behavior measures the learning objective(s). AND Proving behavior uses only general criteria for measuring success.	Proving behavior measures the learning objective(s). AND Proving behavior includes specific criteria (quantitative or qualitative) for measuring success.	*All of level 3 and...* Proving behavior is measured by multiple methods.	• *Is the proving behavior aligned with the learning objective(s)?* • *How is success on the proving behavior measured?*
1.2 Organize instructional plans to promote standards-based, cognitively engaging learning for students	A) Designing and sequencing of learning experiences	The design of the learning experiences is not aligned to the learning objective(s). AND Learning experiences are not sequenced to enable students to demonstrate independent mastery of the learning objective(s) through the gradual release of responsibility.	The design of the learning experiences is not aligned to the learning objective(s). OR Learning experiences are not sequenced to enable students to demonstrate independent mastery of the learning objective(s) through the gradual release of responsibility.	The design of the learning experiences is sequenced to enable students to demonstrate independent mastery of the learning objective(s) through the gradual release of responsibility.	*All of level 3 and...* The design of the learning experiences is differentiated to meet the needs of subgroups of students.	• *Are the learning experiences aligned to the learning objective(s)?* • *Are the learning experiences sequenced using gradual release of responsibility?* • *Level IV: Are learning experiences differentiated?* *NOTE: Instructional pacing is not assessed in this indicator. It is assessed using the phrase 'appropriate time' in indicator 1.2B*
	B) Creating cognitively engaging learning experiences for students	Instructional plans do not provide opportunity for cognitively engaging learning experiences throughout the lesson cycle.	Instructional plans include cognitively engaging learning experiences but without appropriate time and support throughout the lesson cycle.	Instructional plans include cognitively engaging learning experiences throughout the lesson cycle, and each learning experience provides appropriate time and support.	*All of level 3 and...* Instructional plans provide differentiated, cognitively engaging learning experiences for subgroups of students.	• *Are learning experiences consistently cognitively engaging?* • *Does the teacher plan appropriate time and support for students to fully engage in each learning experience?* • *Level IV: Are learning experiences differentiated?*
1.3 Use student data to guide planning	A) Lesson design guided by data	The teacher does not use student data to guide or inform planning.	The teacher uses student data to inform planning of content organization ***or*** instructional strategies. OR The teacher uses student data to inform planning that meets the needs of the whole class.	The teacher uses student data to inform planning of content organization ***and*** instructional strategies. AND The teacher uses student data to inform planning that meets the needs of subgroups of students.	*All of level 3 and...* The teacher cites instructional strategies to meet the needs of individual students.	• *Does the teacher use data to inform content organization and instructional strategies?* • *Is the data used to inform planning for the whole class, subgroups, or individual students?*

Figure 1. Sample organization of the CRTF. *Note.* The table shows the Domain (e.g., Domain 1: Data-Driven Planning and Assessing Student Learning), standards (e.g., Standard 1: Establish standards-based learning objectives and assessments, indicator level (e.g., 1.1A: Selection of learning objectives), and rubric language for each score (e.g., Level 1: Learning objective(s) are missing a specific level of cognition or content).

A teacher's TES is calculated differently based on the group to which a teacher belongs. The TES is a weighted formula consisting of any/all of the following measures: (a) classroom observation score, (b) parent survey score, (c) peer survey score, (d) student survey score, (e) school SGP score, (f) individual SGP score, and (g) SpEd compliance score. Table 1 describes the TES breakdown by teacher group.

Table 1

Teacher Effectiveness Score Breakdowns

Teacher Type	Observation	Parent Survey	Peer Survey	Student Survey	School SGP	Individual SGP	SpEd Compliance
Tested	40%	5%	5%	10%	10%	30%	0%
Non-Tested	55%	5%	5%	10%	25%	0%	0%
SpEd	35%	5%	5%	10%	20%	0%	25%

Performance bands (also known as Effectiveness Bands) denote a teacher's level of effectiveness based on the score he/she has earned. Listed from highest to lowest, the scores are assigned the following labels: Highly Effective 2 (345-400 points), Highly Effective (310-344), Effective (270-309), Emerging (230-269), and Entry (100-229).

Teacher Observation Score (TObs). The TObs makes up the majority of the composite TES. Classroom observations evaluate teachers on indicators for Domains 1, 2, 3, and 4 (4. 1a and 4.1b) of the teaching framework. School administrators, who Green Dot trains and certifies as classroom evaluators through rigorous calibration sessions, observe each teacher a minimum of three times each semester. Two observations are informal, where the observer can drop into the classroom for an unscheduled observation. These observations consist of coaching sessions via a debriefing meeting by the administrator after a short, unannounced observation. Because these sessions are coaching in nature, informal observations are not rated and are non-evaluative towards a teacher's formal observation.

The final observation, known as the formal observation, requires the administrator to observe an entire class period. Formal observations consist of a pre-observation conference where the teacher's lesson plan is evaluated on Domain 1 of the CRTF. The pre-observation conference allows both teacher and evaluator to outline the goals of the formal observation, examine the lesson to be evaluated, evaluate the lesson plan against the CRTF rubric, and discuss any concerns both parties have regarding the upcoming observation. This follows with the classroom observation itself, where Domains 2 and 3 are evaluated. During the classroom observation, the evaluator is physically present during a teacher's instructional time. The evaluator takes in-depth scripting notes of the observed period including instruction, student-to-student interaction, and teacher-to-student interaction, among others. Immediately after the observation, the evaluator will review the collected evidence by tagging the evidence to indicators within the CRTF. Once tagged, both evaluator and teacher can rate the evidence against the CRTF rubric. The final step of the formal observation is the teacher reflection and the collection artifacts such as student work. During this step, the teacher is evaluated on a portion of Domain 4 (indicators 4.1a and 4.1b).

This evaluation cycle occurs once per semester for a total of six observations per year. A weighted average consisting of 40% of a teacher's semester one observation score and 60% of a teacher's semester two observation score make up the teacher's observation score.

Student growth percentile (SGP). A student's growth is calculated using pathways, peer groups, and outcomes. First, a student's pathway is the student's CST testing sequence over a 2-year period. For example, Student A completed the Algebra I CST last year and completes the Algebra II CST during the current year. This student falls under the Algebra I to Algebra II pathway. Next, a student becomes part of a peer group based on his/her starting score in that pathway. In the current example, Student A's starting score in the Algebra I CST is 485. This student's peer group, then, includes all the students in the pathway having the same starting score of 485. Student A will be compared against this group based on his/her performance outcome in the second test. If Student A's outcome is a scale score of 502, a student's individual SGP in the current subject can be determined by comparing his/her outcome with the outcomes of the others in the peer group. This comparison is accomplished using percentile ranking. Students' outcome scores are ranked from 1-100. Where a student ranks in this percentile ranking represents the student's SGP in the subject area.

SGPs can be calculated for the teacher and the school as well. A tested teacher's individual SGP is defined as the median score of all of his/her student's SGPs. A school's overall SGP is the median score of all students in all subject areas.

SGP scores roughly represent student growth in years. An SGP of 50-60 represents growth of approximately 1-1.25 years. SGPs above 50 generally represent growth that exceeds an academic year. Table 2 denotes the SGP designations and what they mean.

Table 2

Student Growth Percentiles: What They Mean

SGP Range	Description	Approximate Growth
75-99	Truly outstanding growth	1.5-2 grade levels
60-75	High growth via excellent teaching	1.25-1.5 grade levels
50-60	Good teaching and good growth	1-1.25 grade levels
40-50	Slightly subpar growth	0.75-1 grade levels
25-40	Subpar growth and/or misalignment with CSTs	0.5-0.75 grade levels
1-25	Low growth and/or misalignment	0.0-0.5 grade levels

Stakeholder feedback. Stakeholder feedback was collected from three separate surveys: 360 surveys, student surveys, and family surveys. Peer surveys measure mastery of Domain 4 of the CRTF and account for five percent of a teacher's TES regardless of the teacher type. Each Green Dot instructor receives anonymous feedback for the indicators in Domain 4 by three peers.

Student surveys are administered twice a year to anywhere from one to 65 randomly select students across all of an instructor's classes. Students are asked to respond to Likert-style questions (e.g., strongly agree, agree, disagree, strongly disagree) about elements of their instructors' teaching practices (Domain 2 and 3 of the CRTF). The student survey represents 10% of the teacher's TES regardless of teacher type. Student responses are tied to their specific instructor.

Parents and/or families complete one survey based on their satisfaction with their child's school learning environment, a measurement of a school's performance on Domain 5 of the CRTF rubric. Families are asked to denote their evaluation using a Likert-style scale. Family responses refer to the school-learning environment as a whole. Therefore, all teachers at the school receive the average score in this measure. This measure comprises five percent of the teacher's TES, regardless of the teacher type.

Restatement of Research Questions

Research questions that guided the study, as identified in Chapter 1, were:

1. Are there differences in student outcomes (e.g., GPA and CST scores) based on a teacher's overall teacher effectiveness score?
2. Are there differences in student outcomes (e.g., GPA, CST, and SGP scores) based on a teacher's overall teacher observation score?
3. Are there differences in student outcomes (e.g., GPA, CST, and SGP scores) based on a teacher's score in a particular CRTF Domain?
4. Are there other observable constructs other than the five Domains identified by CRTF designers that can be arrived at through statistical analysis methods such as factor analysis?
5. Are there differences in student outcomes (e.g., GPA, CST, and SGP scores) based on a teacher's score in a particular CRTF construct arrived at through factor analysis?
6. Are there indicators (or combinations of indicators) that can predict positive student outcomes?

Variables in the Study and Their Levels of Measurement

The following variables were considered in order to address the research questions in the study. These variables, including their corresponding levels of measurement, are listed in Table 3.

Table 3

Variable Table

Variable Name	Level of Measurement
Teacher Effectiveness Score (TES)	Numeric
Teacher Observation Score (TObs)	Numeric
CRTF Domain 1 (D1)	Numeric
CRTF Domain 2 (D2)	Numeric
CRTF Domain 3 (D3)	Numeric
CRTF Domain 4 (D4)	Numeric
CRTF Factor 1 (F1)	Numeric
CRTF Factor 2 (F2)	Numeric
CRTF Factor 3 (F3)	Numeric
CRTF Factor 4 (F4)	Numeric
Student Growth Percentile (SGP)	Numeric
CST Score (CST)	Numeric
GPA Semester 1 (GPAS1)	Numeric
GPA Semester 2 (GPAS2)	Numeric
Individual indicator score (ex. 1_1_A_S1)	Numeric

Statistical Analysis

The NCSS Statistical Analysis suite was used to analyze the quantitative secondary data from Green Dot. Non-nominal demographic information such as school and grade were analyzed using descriptive statistics.

Correlation analysis followed and was used to address research questions 1, 2, 3, and 5. For each of the research questions, a correlation matrix was reported. For questions 3 and 5, regression analysis was used to determine a predictability model for the Domains.

For research question 4, factor analysis was conducted with Varimax rotations. Minimum factor loadings were held at 0.6. Once the Factors were identified, factor scores referred to as CRTF Constructs were calculated. Internal consistency of each factor was reported using Cronbach's alpha.

Finally, for research question number 5, a step-wise regression analysis was conducted to determine a statistically significant regression model that would allow prediction of student outcomes based on the variables measured in the study. For a rubric of this size with 29 indicators, the step-wise regression analysis was deemed the most appropriate since the independent variables are entered according to their statistical contribution in explaining the variance in the dependent variable.

Chapter 4: Results

This chapter presents the results of the study. Correlational analyses were used to study relationship between student outcomes (CST, SGP, and GPA) and TES, TObs, and CRTF Domain scores (D1-D4). The study used archived secondary data collected by Green Dot from the 2012-2013 school year. Tested and non-tested teachers and their students were included in the sample. The section is organized by research question, presenting pertinent data.

Description of the Sample

The sample for this study consisted of 6,604 students. Of these students, 523 (eight percent) were in sixth grade, 544 (nine percent) were in seventh grade, 391 (six percent) were in eighth grade, 1,362 (22%) were in ninth grade, 1,721 (28%) were in 10th grade, and 1,623 (27%) were in 11th grade. Taken together, the sample of students makes up approximately 67% of the student population. There were 295 teachers in the sample. Of the teachers, 257 (87%) were high school teachers (grades nine through 11) and 38 (13%) were middle school teachers (grades six through eight).

Research Question #1

Research question #1 asked, Are there differences in student outcomes (e.g., GPA and CST scores) based on a teacher's overall teacher effectiveness score? To address this research question, correlation analysis was conducted among SGP, student standardized test scores (CST), grade point average for semester one (GPAS1), grade point average for semester two (GPAS2) and the composite TES. A correlation matrix was developed and reported in Table 4.

Table 4

Correlation Matrix for Research Question #1

	CST	SGP	GPAS1	GPAS2	TES
CST	1.000	0.612	0.412	0.428	0.297
SGP		1.000	0.145	0.187	0.292
GPAS1			1.000	0.861	-0.017
GPAS2				1.000	0.034
TES					1.000

Note. All data reported were statistically significant at alpha < 0.05.

As can be seen, there were statistically significant relationships ($p < 0.05$) between Student Achievement Score (CST), SGP, GPAS1, GPAS2, and TES. According to Gertsman (n.d.), Pearson correlation coefficients of $|r| < 0.3$ indicate no correlation, correlations between $0 < |r| < 0.3$ are considered weak, correlations between $0.3 < |r| < 0.7$ are considered moderate, and correlations of $|r| > 0.7$ are considered strong. Accordingly, all three statistically significant correlations were weak. Among them, CST ($r = 0.297$) was the largest positive correlation. By squaring the correlation coefficient, r, one can arrive at the *coefficient of determination* (Gertsman, n.d.; Taylor, 1990). That is, TES only accounted for 8.8% of variation in CST. Correlations between TES and GPAS1 and GPAS2 were not remarkable; essentially no relationship was found between the variables.

To explore other possible relationships between TES and measures of student success, the data were further disaggregated by subject area and the correlation tests run again separately for each. That is, the data were filtered for each subject area separately into a new data set containing only one subject. Each new data set produced a new correlation matrix. Each correlation matrix developed is presented subsequently.

The resulting data set analysis for ELA only was also found to show statistically significant correlations ($p < 0.05$) that were weak ($r < 0.3$) between all variables (CST, SGP, GPAS1, GPAS2, and TES). Among them, CST (*r* = 0.207) was the largest positive correlation showing that TES accounted for merely 4.3% of the variation in CST. SGP (*r* = 0.160) had even weaker correlation. As was the case for all subjects, regarding TES, GPAS1, and GPAS2, no relationships were found between the variables (see Table 5).

Table 5

Correlation Matrix for Research Question #1 Disaggregated (ELA Only)

	CST	SGP	GPAS1	GPAS2	TES
CST	1.000	0.536	0.484	0.499	0.207
SGP		1.000	0.153	0.200	0.160
GPAS1			1.000	0.861	-0.026
GPAS2				1.000	0.037
TES					1.000

Note. All data reported were statistically significant at alpha < 0.05.

When the data were disaggregated for history, the relationships between CST, SGP, GPAS1, GPAS2, and TES were statistically significant. Of the variables, CST (*r* = 0.313) had the largest positive correlation coefficient, with SGP following closely at *r* = 0.285. TES accounted for 9.8% of the variation in CST and only 8.1% of the variation in SGP. Regarding TES, GPAS1, and GPAS2, no relationships were found among the variables (see Table 6).

Table 6

Correlation Matrix for Research Question #1 Disaggregated (History Only)

	CST	SGP	GPAS1	GPAS2	TES
CST	1.000	0.684	0.389	0.409	0.313
SGP		1.000	0.134	0.180	0.285
GPAS1			1.000	0.859	0.018
GPAS2				1.000	0.075
TES					1.000

Note. All data reported were statistically significant at alpha < 0.05.

For math alone, the relationships between CST, SGP, GPAS1, GPAS2, and TES were also statistically significant. Of the variables, CST (*r* = 0.348) scores moderately correlated with

TES. SGP ($r = 0.310$) also moderately correlated with TES. TES accounted for 12.11% of the variation in CST and only 9.6% of the variation in SGP. The relationships among TES, GPAS1, and GPAS2 were again negligible.

Table 7

Correlation Matrix for Research Question #1 Disaggregated (Math Only)

	CST	SGP	GPAS1	GPAS2	TES
CST	1.000	0.618	0.384	0.404	0.348
SGP		1.000	0.106	0.145	0.310
GPAS1			1.000	0.857	-0.084
GPAS2				1.000	-0.055
TES					1.000

Note. All data reported were statistically significant at alpha < 0.05.

In science, all relationships between CST, SGP, GPAS1, GPAS2, and TES were statistically significant. Science yielded the largest correlations thus far in the current study. CST ($r = 0.327$) scores moderately correlated with TES as did SGP ($r = 0.397$). TES accounted for 10.7% of the variation in CST Lastly, TES accounted for 15.8% of the variation in SGP scores. The relationships among TES, GPAS1, and GPAS2 were negligible.

Table 8

Correlation Matrix for Research Question #1 Disaggregated (Science Only)

	CST	SGP	GPAS1	GPAS2	TES
CST	1.000	0.694	0.440	0.455	0.327
SGP		1.000	0.207	0.244	0.397
GPAS1			1.000	0.867	0.046
GPAS2				1.000	0.107
TES					1.000

Note. All data reported were statistically significant at alpha < 0.05.

Research Question #2

Research question #2 asked, Are there differences in student outcomes (e.g., GPA, CST, and SGP scores) based on a teacher's overall teacher observation score? To address this research question, correlation analysis was also conducted among SGP, CST, GPAS1, GPAS2, and TObs. A correlation matrix was developed and reported in Table 9.

Table 9

Correlation Matrix for Research Question #2

	CST	SGP	GPAS1	GPAS2	TObs
CST	1.000	0.612	0.411	0.427	0.162
SGP		1.000	0.147	0.188	0.158
GPAS1			1.000	0.858	-0.025
GPAS2				1.000	0.012
TObs					1.000

Note. All data reported were statistically significant at alpha < 0.05.

As seen in the data, all relationships among CST, SGP, GPAS1, GPAS2, and TObs were statistically significant ($p < 0.05$). All statistically significant correlations were weak. TObs accounted for only 2.6% of the variation in CST and only 2.5% of the variation in SGP. Among them, CST ($r = 0.162$) and SGP ($r = 0.158$) were among the largest correlations. GPAS1 and GPAS2 had no relationship with TObs.

To consider other possible relationships between TObs and measures of student achievement, the dataset was also further disaggregated by subject area and the correlation tests run again separately for each. Each correlation matrix developed for each subject area is reported subsequently.

Disaggregated for science only, all relationships among CST, SGP, GPAS1, GPAS2, and TObs were statistically significant ($p < 0.05$). All correlations related to science were weak. CST ($r = 0.207$) scores and SGP ($r = 0.221$) generated some of the largest correlations in the data set. TObs, then, accounted for only 4.2% of the variation in CST and 4.9% of the variation in SGP. Relationships among TObs, GPAS1, and GPAS2 were again negligible and showed no correlation (see Table 10).

Table 10

Correlation Matrix for Research Question #2 Disaggregated (Science Only)

	CST	SGP	GPAS1	GPAS2	TObs
CST	1.000	0.692	0.437	0.450	0.204
SGP		1.000	0.204	0.237	0.221
GPAS1			1.000	0.867	0.022
GPAS2				1.000	0.070
TObs					1.000

Note. All data reported were statistically significant at alpha < 0.05.

Disaggregated for math only, relationships among CST, SGP, GPAS1, GPAS2, and TObs were statistically significant ($p < 0.05$). As with science, all correlations related to math were weak. CST scores ($r = 0.180$) represented the largest of the correlations in the dataset. This represents TObs accounting for merely 3.2% of the variation in CST scores. TObs and GPAS1 and GPAS2 were not correlated (see Table 11).

Table 11

Correlation Matrix for Research Question #2 Disaggregated (Math Only)

	CST	SGP	GPAS1	GPAS2	TObs
CST	1.000	0.617	0.391	0.413	0.180
SGP		1.000	0.122	0.158	0.158
GPAS1			1.000	0.850	-0.085
GPAS2				1.000	-0.065
TObs					1.000

Note. All data reported were statistically significant at alpha < 0.05.

In history only, all relationships among CST, SGP, GPAS1, GPAS2, and TObs had statistical significance ($p < 0.05$). All correlations were weak. CST scores (r = 0.152) and SGP scores ($r = 0.137$) represented the largest of the correlations in the dataset, accounting for only 2.3% and 1.9%, respectively, of the variation in TObs. TObs, GPAS1, and GPAS2 were again not found to be correlated (see Table 12).

Table 12

Correlation Matrix for Research Question #2 Disaggregated (History Only)

	CST	SGP	GPAS1	GPAS2	TObs
CST	1.000	0.682	0.370	0.389	0.152
SGP		1.000	0.127	0.173	0.137
GPAS1			1.000	0.856	-0.014
GPAS2				1.000	0.037
TObs					1.000

Note. All data reported were statistically significant at alpha < 0.05.

Looking at ELA only, all relationships were statistically significant. Among the relationships, CST and SGP were both weak, yet showed the largest correlations with $r = 0.132$ and $r = 0.124$, respectively. TObs only accounted for 1.7% of the variation in CST and only 1.5% of the variation in SGP. The Pearson correlation among TObs, GPAS1, and GPAS2 was also negligible and therefore had no relationship (see Table 13).

Table 13

Correlation Matrix for Research Question #2 Disaggregated (ELA Only)

	CST	SGP	GPAS1	GPAS2	TObs
CST	1.000	0.538	0.487	0.502	0.132
SGP		1.000	0.158	0.204	0.124
GPAS1			1.000	0.858	-0.031
GPAS2				1.000	0.001
TObs					1.000

Note. All data reported were statistically significant at alpha < 0.05.

Research Question #3

Research question #3 asked, Are there differences in student outcomes (e.g., GPA, CST, and SGP scores) based on a teacher's score in a particular CRTF Domain? Correlation analysis was also employed to determine the relationships between SGP, CST, GPAS1, GPAS2, Domain 1 of the CRTF (D1S1), Domain 2 of the CRTF (D2S1), Domain 3 of the CRTF (D3S1), and

Domain 4 of the CRTF (D4S1). A correlation matrix was developed and is reported subsequently.

Green Dot offers teachers the opportunity to transfer a score of a 3 or a 4 on any indicator to the second semester formal evaluation. Because domain scores in semester two formal evaluations may be tainted by this protocol, domain scores were taken from domain averages during semester one only (see Table 14).

Table 14

Correlation Matrix for Research Question #3 - All Subjects

	CST	SGP	GPAS1	GPAS2	D1S1	D2S1	D3S1	D4S1
CST	1.000	0.612	0.413	0.428	0.104	0.189	0.140	0.069
SGP		1.000	0.146	0.188	0.123	0.163	0.153	0.084
GPAS1			1.000	0.860	-0.034	0.004	-0.002	0.037
GPAS2				1.000	-0.000	0.016	0.012	0.056
D1S1					1.000	0.522	0.582	0.455
D2S1						1.000	0.771	0.331
D3S1							1.000	0.431
D4S1								1.000

Note. All data reported were statistically significant at alpha < 0.05.

As can be seen, there were statistically significant relationships among CST, SGP, GPAS1, GPAS2, D1S1, D2S1, D3S1, and D4S1. According to Gertsman's (n.d.) interpretation of the Pearson coefficient, all relationships were reported as weak. Among the largest of the correlations was found between Domain 2 and CST ($r = 0.189$) and SGP ($r = 0.163$). Domain 2 accounted for 3.6% of the variation in CST and 2.7% of the variation in SGP. Relationships among all Domains and GPAS1 and GPAS2 were less than $r = 0.10$ and therefore indicated no relationship. Additionally, there was no relationship among Domain 4, CST, and SGP.

Similarly to research questions #1 and #2, the dataset was also further disaggregated by subject area to examine other relationships among the CRTF Domains and measures of student achievement. The resulting filtered dataset was used for correlation. Each correlation matrix developed for each subject area is presented subsequently.

Disaggregated for science only, all relationships among CST, SGP, GPAS1, GPAS2, D1S1, D2S1, D3S1, and D4S1 were statistically significant ($p < 0.05$). All relationships exhibited a weak to nonexistent correlation. Among them, Domain 2 and Domain 3 had the largest Pearson coefficients between CST ($r = 0.256$ for Domain 2 and $r = 0.226$ for Domain 3) and SGP ($r = 0.263$ for Domain 2 and $r = 0.279$ for Domain 3). That is, Domain 2 accounted for 6.6% of the variation in CST and 6.9% of the variation in SGP. Domain 3 accounted for only 5.1% of the variation in CST and 7.8% of the variation in SGP (see Table 15).

Table 15

Correlation Matrix for Research Question #3 - Disaggregated (Science Only)

	CST	SGP	GPAS1	GPAS2	D1S1	D2S1	D3S1	D4S1
CST	1.000	0.694	0.440	0.455	0.141	0.256	0.226	0.217
SGP		1.000	0.207	0.244	0.141	0.263	0.279	0.167
GPAS1			1.000	0.867	0.080	0.006	0.037	0.030
GPAS2				1.000	0.112	0.043	0.080	0.077
D1S1					1.000	0.631	0.649	0.616
D2S1						1.000	0.819	0.549
D3S1							1.000	0.562
D4S1								1.000

Note. All data reported were statistically significant at alpha < 0.05.

Exploring math alone, all relationships with CST, SGP, GPAS1, GPAS2, D1S1, D2S1, D3S1, and D4S1 were statistically significant ($p < 0.05$). All correlations exhibited a weak to nonexistent correlation. The largest correlations, though weak, were found in Domain 2. CST scores ($r = 0.217$) and SGP scores ($r = 0.169$) accounted for only 4.7% and 2.9%, respectively, of the variation in Domain 2. Domain 4 had the smallest correlations ($r < 0.1$) and therefore had no relationships with any student achievement measure. GPAS1 and GPAS2 both had weak negative correlations with Domain 1, but showed no relationships with the other Domains (see Table 16).

Table 16

Correlation Matrix for Research Question #3 - Disaggregated (Math Only)

	CST	SGP	GPAS1	GPAS2	D1S1	D2S1	D3S1	D4S1
CST	1.000	0.620	0.384	0.402	0.150	0.217	0.151	-0.021
SGP		1.000	0.114	0.148	0.151	0.169	0.162	0.010
GPAS1			1.000	0.855	-0.127	0.010	-0.013	0.039
GPAS2				1.000	-0.099	0.019	-0.012	0.015
D1S1					1.000	0.585	0.718	0.253
D2S1						1.000	0.752	0.253
D3S1							1.000	0.345
D4S1								1.000

Note. All data reported were statistically significant at alpha < 0.05.

Examining history only, all relationships with CST, SGP, GPAS1, GPAS2, D1S1, D2S1, D3S1, and D4S1 were statistically significant ($p < 0.05$). All correlations in this case exhibited a weak to nonexistent correlation. The largest correlations were reported in Domain 4. Domain 4 accounted for only two percent of the variation in CST scores ($r = 0.140$) and only 1.9% of the variation in SGP scores ($r = 0.139$). GPAS1 and GPAS2 had the smallest correlations for all Domains ($r < 0.1$) except Domain 4, where GPAS2 yielded $r = 0.112$ (see Table 17).

Table 17

Correlation Matrix for Research Question #3 - Disaggregated (History Only)

	CST	SGP	GPAS1	GPAS2	D1S1	D2S1	D3S1	D4S1
CST	1.000	0.684	0.389	0.409	0.113	0.154	0.089	0.140
SGP		1.000	0.134	0.180	0.097	0.137	0.120	0.139
GPAS1			1.000	0.859	-0.074	0.026	-0.020	0.077
GPAS2				1.000	0.001	0.044	0.000	0.112
D1S1					1.000	0.409	0.588	0.533
D2S1						1.000	0.751	0.180
D3S1							1.000	0.357
D4S1								1.000

Note. All data reported were statistically significant at alpha < 0.05.

Disaggregating for ELA only, all relationships with CST, SGP, GPAS1, GPAS2, D1S1, D2S1, D3S1, and D4S1 were also statistically significant ($p < 0.05$). All correlations in this case exhibited a weak to nonexistent correlation, with most relationships having an r-value of less than 0.1. The largest correlations were reported between CST and Domain 2 ($r = 0.139$) and Domain 3 ($r = 0.116$). CST score accounted for only 1.9% of the variation in Domain 2 and only 1.3% of the variation in Domain 3. GPAS1 and GPAS2 yielded very small negative correlations for all Domains ($r < 0.1$) except Domain 4 (see Table 18).

Table 18

Correlation Matrix for Research Question #3 Disaggregated (ELA Only)

	CST	SGP	GPAS1	GPAS2	D1S1	D2S1	D3S1	D4S1
CST	1.000	0.534	0.484	0.500	0.055	0.139	0.116	0.054
SGP		1.000	0.151	0.198	0.107	0.082	0.081	0.076
GPAS1			1.000	0.859	-0.039	-0.015	-0.019	0.006
GPAS2				1.000	-0.021	-0.026	-0.021	0.038
D1S1					1.000	0.475	0.432	0.506
D2S1						1.000	0.782	0.414
D3S1							1.000	0.455
D4S1								1.000

Note. All data reported were statistically significant at alpha < 0.05.

Multiple regression analysis for question #3 (all subjects). Multiple regression analysis was employed for all subjects to predict student outcomes through a regression model from all Domains (see Table 19).

SGP. For the first iteration, all variables were statistically significant ($p < 0.05$), with the exception of Domain 4 ($p = 0.157$). Domain 4 was removed and the model was run again. At the end of this iteration, all variables were statistically significant, leaving a regression model involving Domains 1-3. This model accounted for 2.9% of the variability in SGP.

Table 19

Multiple Regression Summary Table (All Subjects) – Showing Final Iterations

	SGP			CST			GPAS2			GPAS1		
IV	B	Beta	P	B	Beta	P	B	Beta	P	B	Beta	P
D1S1	2.29	0.04	0.00	*	*	*	-0.06	-0.03	0.00	-0.13	-0.07	0.00
D2S1	6.92	0.10	0.00	24.6	0.19	0.00	0.11	0.07	0.00	0.04	0.02	0.00
D3S1	3.59	0.05	0.00	*	*	*	*	*	*	*	*	*
D4S1	*	*	*	*	*	*	*	*	*	0.09	0.06	0.00
R^2	0.0293			0.0354			0.004			0.0051		
MODEL	17.83+2.29(D1S1)+6.92(D2S1)+ 3.59(D3S1)			257.66+24.612(D2S1)			2.24-0.059(D1S1)+0.111(D2S1)			2.36- 0.12(D1S1)+0.043(D2S1)+0.095(D4S1)		

* Variable rejected from the final model during previous iterations (*p* > 0.05)

CST. Domains 4, 1, and 3 were removed in this order in subsequent iterations due to *p* values > 0.05. As a result, Domain 2 remained the only variable in a regression model accounting for 3.5% of the variability in CST.

GPA. A regression model involving D1S1 and D2S1 accounted for only 0.4% of the variability in GPAS1. Another model involving D1S1, D2S1, and D4S1 accounted for 1.5% of the variability in GPAS2.

Multiple regression analysis (disaggregated for different subjects). The same multiple regression analyses were employed again for each of the different subject areas to determine if similar results are observed for different subjects (Tables 20-23). All statistically insignificant ($p > 0.05$) variables were removed from subsequent runs and then the model was run again.

SGP. For ELA, the regression model involving Domains 1 and 3 accounted for 1.3% ($R^2 = 0.013$) of the variation seen in SGP. For history, the regression model involving Domains 2 and 4 accounted for only 3.2% ($R^2 = 0.032$) of the variation in SGP. For math, the regression model involving Domains 1-4 accounted for 3.5% ($R^2 = 0.035$) of the variation in SGP. For science, the model incorporating Domains 1-3 accounted for 8.5% ($R^2 = 0.085$) of the variation in SGP.

CST. For ELA, only Domain 2 had statistical significance, accounting for merely 1.9% ($R^2 = 0.019$) of the variation in CST. In history, the model involving Domains 2-4 accounted for 4.4% ($R^2 = 0.044$) of the variation in CST. Math yielded a higher R^2 at 0.053; Domains 2, 3 and 4 accounted for 5.3% of the variation in CST. However, as in SGP, the model for science (Domains 1, 2, and 4) accounted for the most variation in CST score, 7.8%.

GPA. For semester two ELA, Domains 1, 3 and 4 accounted for 0.5% (R^2 = 0.005) of the variation in GPA. History (Domains 1-4), math (Domains 1, 2, and 4), and science (Domains 1, 2, and 3) accounted for 2.4% (R^2 = 0.024), 2% (R^2 = 0.020), and 1.6% (R^2 = 0.016), respectively, of the variation in GPA for semester two.

For semester one ELA only Domain 1 had statistical significance accounting for 0.2% (R^2 = 0.002) of the variation in GPA. History (Domains 1-4), math (Domains 1-4), and science (Domains 1 and 2), accounted for 3.1% (R^2 = 0.031), 3.3% (R^2 = 0.033), and 1% (R^2 = 0.0098), respectively, of the variation in GPA for semester one.

Table 20

Multiple Regression Summary Table (Disaggregated for ELA Only) – Showing Final Iterations

	SGP			CST			GPAS2			GPAS1		
IV	B	Beta	P	B	Beta	P	B	Beta	P	B	Beta	P
D1S1	5.39	0.09	0.00	*	*	*	-0.09	-0.04	0.01	-0.10	-0.05	0.00
D2S1	*	*	*	17.57	0.14	0.00	*	*	*	*	*	*
D3S1	3.15	0.04	0.00	*	*	*	-0.09	-0.04	0.02	*	*	*
D4S1	*	*	*	*	*	*	0.14	0.08	0.00	0.06	0.03	0.04
R^2	0.013			0.019			0.005			0.002		
MODEL	28.99+5.39(D1S1)+3.15(D3S1)			280.32+17.57(DS21)			2.45-0.09(D1S1)-0.087(D2S1)+0.14(D4S1)			2.52-0.10(D1S1)+0.06(D4S1)		

* Variable rejected from the final model during previous iterations ($p > 0.05$)

Table 21

Multiple Regression Summary Table (Disaggregated for History Only) – Showing Final Iterations

	SGP			CST			GPAS2			GPAS1		
IV	B	Beta	P	B	Beta	P	B	Beta	P	B	Beta	P
D1S1	*	*	*	*	*	*	-0.14	-0.08	0.00	-0.31	-0.18	0.00
D2S1	7.24	0.12	0.00	30.65	0.23	0.00	0.27	0.13	0.00	0.22	0.12	0.00
D3S1	*	*	*	-21.63	-0.13	0.00	-0.28	-0.11	0.00	-0.15	-0.06	0.03
D4S1	6.68	0.12	0.00	18.11	0.15	0.00	0.32	0.17	0.00	0.31	0.18	0.00
R^2		0.032			0.044			0.024			0.031	
MODEL	18.26+7.24(D2S1)+6.68(D4S1)			247.05+30.66(D2S1)-21.63(D3S1)+18.11(D4S1)			1.7-0.14(D1S1)+0.27(D2S1)-0.28(D3S1)+0.32(D4S1)			1.99-0.31(D1S1)+0.22(D2S1)-0.15(D3S1)+0.31(D4S1)		

* Variable rejected from the final model during previous iterations ($p > 0.05$)

Table 22

Multiple Regression Summary Table (Disaggregated for Math Only) – Showing Final Iterations

	SGP			CST			GPAS2			GPAS1		
IV	B	Beta	P	B	Beta	P	B	Beta	P	B	Beta	P
D1S1	3.75	0.06	0.00	6.69	0.05	0.00	-0.31	-0.17	0.00	-0.42	-0.25	0.00
D2S1	5.95	0.09	0.00	27.30	0.20	0.00	0.22	0.11	0.00	0.15	0.08	0.00
D3S1	4.45	0.06	0.02	*	*	*	*	*	*	0.18	0.09	0.00
D4S1	-1.98	-0.05	0.00	-6.95	-0.07	0.00	0.05	0.04	0.01	0.07	0.06	0.00
R^2		0.035			0.053			0.020			0.033	
MODEL	20.89+3.75(D1S1)+5.95(D2S1)+4.45(D3S1)-1.98(D4S1)			240.37+6.69(D1S1)27.30(D2S1)-6.95(D4S1)			2.57-0.31(D1S1)+0.22(D2S1)+0.05(D4S1)			2.56-0.42(D1S1)+0.15(D2S1)+0.18(D3S1)+0.07(D4S1)		

* Variable rejected from the final model during previous iterations (*p* > 0.05)

Table 23

Multiple Regression Summary Table (Disaggregated for Science Only) – Showing Final Iterations

	SGP			CST			GPAS2			GPAS1		
IV	B	Beta	P	B	Beta	P	B	Beta	P	B	Beta	P
D1S1	-4.97	-0.09	0.00	-9.71	-0.10	0.00	0.21	0.12	0.00	0.20	0.13	0.00
D2S1	8.96	0.13	0.00	28.06	0.24	0.00	-0.22	-0.10	0.00	-0.15	-0.07	0.00
D3S1	15.59	0.23	0.00	*	*	*	0.19	0.09	0.00	*	*	*
D4S1	*	*	*	12.07	0.15	0.00	*	*	*	*	*	*
R^2	0.085			0.078			0.016			0.0098		
MODEL	-1.38-4.97(D1S1)+8.96(D2S1) +15.59(D3S1)			245.15-9.71(D1S1) +28.06(D2S1)+12.07(D4S1)			1.96+0.21(D1S1)- 0.22(D2S1)+0.19(D3S1)			2.29+0.20(D1S1)-0.15(D2S1)		

* Variable rejected from the final model during previous iterations ($p > 0.05$)

Research Question #4

Research question #4 asked, Are there other observable constructs other than the five Domains identified by CRTF designers that can be arrived at through statistical analysis methods such as factor analysis? For question #4, factor analysis was employed to determine if additional Factors or Domains could be observed from the data.

Factor analysis was conducted with Varimax rotation. Minimal factor loadings were set at 0.4. Three, four, five, and six factor solutions were examined. Among them, the five-factor solution accounted for 99.80% of the variance in the model and was selected as the optimal solution.

During the analysis, two indicators were eliminated because they did not meet minimum criteria for primary factor loadings of 0.4 or above. Indicators 3.1A ("Communication of the learning objectives of the lesson") and 3.3D ("Resources and instructional materials") did not load above 0.4 on any factor.

Additionally, four indicators cross-loaded across two separate Factors. Indicator 2.1A ("Value of effort and challenge") loaded across both Factors one and four with loadings on both Factors greater than 0.5. The indicator was eliminated from the factor with the smallest factor loading value: factor four. Indicator 3.3B ("Academic discourse") loaded across Factors one and four as well. The indicator was eliminated from factor one, the factor with the lowest loading value. Indicator 3.2B ("Cognitive level of student learning experiences") cross-loaded on Factors one and four. The indicator was removed from factor one due to low loading values. Finally, indicator 1.3A cross-loaded on both factor 2 and factor 5. Factor 5 had the lower factor loading compared to factor 2; therefore the indicator was eliminated from the factor. Factor 5 only had one factor that initially loaded. By removing this indicator from factor 5, the analysis results ostensibly now show a four-factor solution (see Table 24).

Table 24

Factor Loadings and Communalities Based on Factor Analysis with Varimax Rotation for 29 Items from Green Dot Public School's CRTF

Indicator	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Commonality
1.1A		0.524				0.370
1.1B		0.607				
1.2A		0.726				
1.2B		0.693				
1.3A		0.527			~~0.483~~	
1.4A		0.610				
1.4B		0.589				
1.5A		0.685				
1.5B		0.684				
2.1A	-0.545			~~-0.500~~		
2.2A	-0.732					
2.2B	-0.637					
2.3A	-0.596					
2.3B	-0.650					
2.4A	-0.570					
3.1B				-0.517		0.351
3.1C				-0.534		
3.2A				-0.454		0.384
3.2B	~~-0.404~~			-0.460		
3.3A				-0.549		
3.3B	~~-0.450~~			-0.591		
3.3C	-0.614					
3.4A				-0.580		
3.4B				-0.631		
3.4C				-0.498		
4.1A			0.795			
4.1B			0.658			

Note. Factor loadings < 0.4 are suppressed. For Factors that cross-loaded, the factor with the lowest loading value presented is struck out.

Internal consistency for each of the Factors was examined using Cronbach's alpha. A rule of thumb for interpreting internal consistency, according to George and Mallery (2003), has traditionally been alpha > 0.90 is excellent, 0.90 < *a* < 0.80 is good, 0.80 < *a* < 0.70 is acceptable, 0.70 < *a* < 0.60 is questionable, 0.60 < *a* < 0.50 is poor, and alphas < 0.50 are unacceptable. All alphas were generally good to excellent: 0.872 for factor 1, 0.875 for factor 2, 0.741 for factor 3, and 0.865 for factor 4.

Overall, these analyses indicate that four additional Factors underlie scoring on the CRTF. It is important to note, however, that many of these Factors seem to coincide closely with the Domains already defined by the CRTF designers. Factor 2 and Factor 3 are 100% identical to Domain 1 and Domain 4, respectively. Furthermore, Factors 1 and 4 bore similarities to Domains 2-3, respectively. This is illustrated in Figure 2.

Indicator	Original CRTF	Indicator Text	Factor Analysis
1.1.A		Selection of learning objectives	
1.1.B		Measurability of learning objectives	
1.2.A		Designing and sequencing of learning experiences	
1.2.B		Creating cognitively engaging learning experiences for students	Factor 2
1.3.A	Domain 1	Lesson design guided by data	
1.4.A		Knowledge of subject matter to identify pre-requisite knowledge	
1.4.B		Addresses common content misconceptions	
1.5.A		Selection of assessments	
1.5.B		Planned response to assessment	
2.1.A		Value of effort and challenge	
2.2.A		Behavioral expectations	
2.2.B		Response to behavior	
2.3.A	Domain 2	Interactions between teacher and students	Factor 1
2.3.B		Student interactions with each other	
2.4.A		Routines procedures and transitions	
3.1.A		Communication of the learning objectives of the lesson	
3.1.B		Connections to prior and future learning experiences	
3.1.C		Criteria for success	
3.2.A		Executes lesson cycle	Factor 4
3.2.B		Cognitive Level of Student Learning Experiences	
3.3.A	Domain 3	Questioning	
3.3.B		Academic Discourse	
3.3.C		Group structures	
3.3.D		Resources and instructional materials	
3.4.A		Checking for understanding and adjusting instruction	
3.4.B		Feedback to students	
3.4.C		Self-monitoring	
4.1.A	Domain 4	Accuracy	Factor 3
4.1.B		Use in future planning	

Figure 2. Results of factor analysis showing four new Factors compared to original CRTF Domains.

Research Question #5

Research question #5 asked, Are there differences in student outcomes (e.g., GPA, CST, and SGP scores) based on a teacher's score in a particular CRTF construct arrived at through factor analysis? Correlation analysis was also conducted to determine the r-values between SGP, student standardized test scores (CST), GPAS1, GPAS2, factor 1 (FACT1), factor 2 (FACT2), factor 3 (FACT3), and factor 4 (FACT4). A correlation matrix was developed and is reported in Table 25.

Table 25

Correlation Matrix for New Factors (All Subjects)

	CST	SGP	GPAS1	GPAS2	FACT1	FACT2	FACT3	FACT4
CST	1.000	0.612	0.413	0.428	0.187	0.104	0.069	0.133
SGP		1.000	0.146	0.188	0.163	0.123	0.084	0.153
GPAS1			1.000	0.860	0.016	-0.034	0.037	-0.001
GPAS2				1.000	0.029	-0.000	0.056	0.007
FACT1					1.000	0.537	0.359	0.739
FACT2						1.000	0.455	0.561
FACT3							1.000	0.412
FACT4								1.000

Note. All data reported were statistically significant at alpha < 0.05.

For all subjects, all relationships among CST, SGP, GPAS1, GPAS2, factor 1 (FACT1), factor 2 (FACT2), factor 3 (FACT3), and factor 4 (FACT4) were statistically significant ($p < 0.05$). All statistically significant correlations reported weak to no relationships. Among them, Factor 1 had the largest Pearson coefficient at $r = 0.187$ for CST and $r = 0.163$ for SGP. Factor 1 accounted for only 3.5% of the variation in CST and just 2.7% of the variation in SGP. GPAS1 and GPAS2 had no relationship with other variables.

The dataset was further disaggregated by subject area and the correlation tests run again separately for each in order to explore other relationships. The correlation matrix developed for each subject area is reported subsequently.

Filtering for ELA only, all relationships (CST, SGP, GPAS1, GPAS2, FACT1, FACT2, FACT3, and FACT4) were statistically significant. However, all reported Pearson correlations represented a weak to no relationship. Among them, Factor 1 and CST had the largest correlation at only r = 0.137, accounting for 1.9% of the variation (see Table 26).

Table 26

Correlation Matrix for Factors (ELA Only)

	CST	SGP	GPAS1	GPAS2	FACT1	FACT2	FACT3	FACT4
CST	1.000	0.534	0.484	0.500	0.137	0.055	0.054	0.114
SGP		1.000	0.151	0.198	0.075	0.107	0.076	0.083
GPAS1			1.000	0.859	-0.008	-0.039	0.006	-0.025
GPAS2				1.000	-0.021	-0.021	0.038	-0.017
FACT1					1.000	0.474	0.431	0.757
FACT2						1.000	0.506	0.409
FACT3							1.000	0.424
FACT4								1.000

Note. All data reported were statistically significant at alpha < 0.05.

Looking at history only, all relationships among CST, SGP, GPAS1, GPAS2, FACT1, FACT2, FACT3, and FACT4 were statistically significant. All relationships were weak to non-existent. Among the variables, Factor 1 yielded the highest Pearson coefficients between CST (r = 0.157) and SGP (r = 0.148). That is, factor 1 accounted for only 2.5% and 2.2%, respectively, of the variation in CST and SGP. Factor 3 yielded Pearson coefficients of r = 0.140 for CST and r = 0.139 for SGP. Relationships among GPAS1, GPAS2, FACT1, FACT2, FACT3, and FACT4 were non-existent (see Table 27).

Table 27

Correlation Matrix for Factors (History Only)

	CST	SGP	GPAS1	GPAS2	FACT1	FACT2	FACT3	FACT4
CST	1.000	0.684	0.389	0.409	0.157	0.113	0.140	0.072
SGP		1.000	0.134	0.180	0.148	0.097	0.139	0.103
GPAS1			1.000	0.859	0.037	-0.074	0.077	-0.028
GPAS2				1.000	0.059	0.001	0.112	-0.007
FACT1					1.000	0.445	0.241	0.699
FACT2						1.000	0.533	0.558
FACT3							1.000	0.358
FACT4								1.000

Note. All data reported were statistically significant at alpha < 0.05.

Disaggregating for math only, CST, SGP, GPAS1, GPAS2, FACT1, FACT2, FACT3, and FACT4 yielded relationships that were statistically significant. All relationships were weak to non-existent. Factor 1 again reported the highest Pearson coefficients between CST (r = 0.206) and SGP (r = 0.161), accounting for 4.2% and 2.6%, respectively, of the variation in CST and SGP. Factor 2 and factor 4 had similar relationships with CST and SGP. Factor 3, however, showed no relationship between CST, SGP, GPAS1, and GPAS2. Relationships among GPAS1, GPAS2, FACT1, FACT2, FACT3, and FACT4 were non-existent (see Table 28).

Table 28

Correlation Matrix for Factors (Math Only)

	CST	SGP	GPAS1	GPAS2	FACT1	FACT2	FACT3	FACT4
CST	1.000	0.620	0.384	0.402	0.206	0.150	-0.021	0.160
SGP		1.000	0.114	0.148	0.161	0.151	0.010	0.174
GPAS1			1.000	0.855	0.020	-0.127	0.039	-0.029
GPAS2				1.000	0.032	-0.099	0.015	-0.027
FACT1					1.000	0.593	0.221	0.735
FACT2						1.000	0.253	0.712
FACT3							1.000	0.328
FACT4								1.000

Note. All data reported were statistically significant at alpha < 0.05.

Disaggregating for science only, all relationships among all variables (CST, SGP, GPAS1, GPAS2, FACT1, FACT2, FACT3, and FACT4) were statistically significant, though generally weak. All Factors correlated weakly with CST and SGP. Among the variables, SGP had the largest correlation *r*-values for factor 1 ($r = 0.271$) and factor 4 ($r = 0.273$). That is, factor 1 accounted for 7.3% of the variation and factor 4 accounted for 7.4% of the variation. All Factors had *r*-values greater than 0.1 for both CST and SGP. Relationships among GPAS1, GPAS2, FACT1, FACT2, FACT3, and FACT4 were again negligible (see Table 29).

Table 29

Correlation Matrix for Factors (Science Only)

	CST	SGP	GPAS1	GPAS2	FACT1	FACT2	FACT3	FACT4
CST	1.000	0.694	0.440	0.455	0.261	0.141	0.217	0.202
SGP		1.000	0.207	0.244	0.271	0.141	0.167	0.273
GPAS1			1.000	0.867	0.023	0.080	0.030	0.034
GPAS2				1.000	0.064	0.112	0.077	0.076
FACT1					1.000	0.064	0.564	0.773
FACT2						1.000	0.616	0.628
FACT3							1.000	0.534
FACT4								1.000

Note. All data reported were statistically significant at alpha < 0.05.

Multiple regression analysis for new Factors (all subjects). Multiple regression analysis was employed for all subjects to predict student outcomes through a regression model from the new Factors arrived at from factor analysis (see Table 30).

Table 30

Multiple Regression Summary Table for New Factors (All Subjects) – Showing Final Iterations

	SGP			CST			GPAS2			GPAS1		
IV	B	Beta	P	B	Beta	P	B	Beta	P	B	Beta	P
Factor 1	6.67	0.10	0.00	24.20	0.19	0.00	0.11	0.05	0.00	0.122	0.06	0.00
Factor 2	2.12	0.04	0.00	*	*	*	-0.07	-0.04	0.00	-0.13	-0.07	0.00
Factor 3	*	*	*	*	*	*	0.11	0.07	0.00	0.10	0.07	0.00
Factor 4	3.72	0.06	0.00	*	*	*	-0.08	-0.04	0.00	-0.08	-0.04	0.00
R^2	0.0296			0.0349			0.0054			0.0064		
MODEL	19.14+ 6.67*FACT1+ 2.11*FACT2+ 3.72*FACT4			259.68+ 24.20*FACT1			2.16+ 0.11*FACT1-0.07*FACT2+ 0.11*FACT3-0.08*FACT4			2.33+ 0.12*FACT1-0.13*FACT2+ 0.10*FACT3-0.08*FACT4		

* Variable rejected from the final model during previous iterations ($p > 0.05$)

The regression model involving Factors 1, 2, and 4 accounted for 2.96% ($R^2 = 0.0296$) of the variation in SGP. For CST, only Factor 1 had a statistically significant *p*-value ($p < 0.05$) and accounted for only 1.87% ($R^2 = 0.0187$) of the variation in the model. Factors 1, 2, and 3 were included in a model that accounted for only 0.45% ($R^2 = 0.0045$) of the variation in GPA for semester two. Only Factor 4 was statistically significant in the regression model for GPA semester one, accounting for only 0.07% ($R^2 = 0.0007$) of the variation.

Multiple regression analysis for new Factors (disaggregated for different subjects). To determine if this is the case for different subject areas, the data set was filtered out by subject and the regression analysis was run again (Tables 31-34).

SGP. For ELA, Factors 2 and 4 accounted for 1.31% ($R^2 = 0.0131$) of the variation. For history, the regression model involving Factors 1 and 3 accounted for 3.31% ($R^2 = 0.0331$) of the variation in SGP. All Factors were statistically significant in the regression model for math, accounting for 3.61% ($R^2 = 0.0361$) of the variation in SGP. The regression model for science, involving Factor 1, 2, and 4, however, accounted for the largest variation in SGP: 8.74% ($R^2 = 0.0874$)

CST. Factor 1 was the only statistically significant factor in the model for ELA that accounted for 1.31% ($R^2 = 0.0131$) of the variation in CST. For history, the regression model involving Factors 1, 3, and 4 accounted for 4.28% ($R^2 = 0.0428$) of the variation in CST. For history and science, Factors 1-3 accounted for 4.92% ($R^2 = 0.0492$) and 8.07% ($R^2 = 0.0807$) of the variation in CST, respectively.

GPA. For ELA, the model involving Factors 1-3 accounted for 0.45% ($R^2 = 0.0045$) of the variation in ELA. All Factors were statistically significant in the regression model for history, accounting for 2.75% ($R^2 = 0.0275$). For math, the model that included Factors 1, 2, and 3 accounted for 2.33% ($R^2 = 0.0233$) of the variation in GPA. In science, Factors 1 and 2 only accounted for 0.78% ($R^2 = 0.0078$) of the variation in GPA.

For semester one, Factor 4 was the only significant variable in the model that accounted for 0.07% ($R^2 = 0.0007$) of the variation in GPA for ELA. For history, the model involved all Factors and accounted for 3.41% ($R^2 = 0.0341$) of the variation in GPA. For math, the model involving Factors 1, 2, and 3 can explain 3.37% of the variation in GPA. For science, Factor 1 accounted for just 0.41% ($R^2 = 0.0041$) of the variation in GPA.

Table 31

Multiple Regression Summary Table for New Factors (ELA Only) – Showing Final Iterations

	SGP			CST			GPAS2			GPAS1		
	B	Beta	P	B	Beta	P	B	Beta	P	B	Beta	P
Factor 1	*	*	*	17.43	0.14	0.00	-0.08	-0.03	0.04	*	*	*
Factor 2	5.36	0.09	0.00	*	*	*	-0.08	-0.04	0.01	*	*	*
Factor 3	*	*	*	*	*	*	0.14	0.07	0.00	*	*	*
Factor 4	3.02	0.05	0.00	*	*	*	*	*	*	-0.05	-0.03	0.00
R^2	0.0131			0.0187			0.0045			0.0007		
MODEL	29.55+ 5.36*FACT2+ 3.02*FACT4			281.36+ 17.43*FACT1			2.46-0.08*FACT1-0.08*FACT2+ 0.14*FACT3			2.56-0.05*FACT4		

* Variable rejected from the final model during previous iterations (*p* > 0.05)

Table 32

Multiple Regression Summary Table for New Factors (History Only) – Showing Final Iterations

	SGP			CST			GPAS2			GPAS1		
	B	Beta	P	B	Beta	P	B	Beta	P	B	Beta	P
Factor 1	7.57	0.12	0.00	28.02	0.21	0.00	0.28	0.14	0.00	0.25	0.13	0.00
Factor 2	*	*	*	*	*	*	-0.16	-0.08	0.00	-0.31	-0.18	0.01
Factor 3	6.19	0.11	0.00	16.40	0.13	0.00	0.31	0.17	0.01	0.30	0.17	0.01
Factor 4	*	*	*	-17.74	-0.12	0.00	-0.26	-0.12	0.00	-0.17	-0.08	0.00
R^2	0.0331			0.0428			0.0275			0.0341		
MODEL	19.03+ 7.57*FACT1+ 6.19*FACT3			250.05+ 28.02*FACT1+ 16.40*FACT3-17.74*FACT4			1.67+ 0.28*FACT1-0.16*FACT2+ 0.31*FACT3-0.26*FACT4			1.98+ 0.25*FACT1-0.31*FACT2+ 0.30*FACT3-0.17*FACT4		

* Variable rejected from the final model during previous iterations ($p > 0.05$)

Table 33

Multiple Regression Summary Table for New Factors (Math Only) – Showing Final Iterations

	SGP			CST			GPAS2			GPAS1		
	B	Beta	P	B	Beta	P	B	Beta	P	B	Beta	P
Factor 1	3.90	0.06	0.00	25.32	0.19	0.00	0.26	0.14	0.00	0.25	0.14	0.00
Factor 2	2.92	0.05	0.02	7.55	0.06	0.00	-0.33	-0.19	0.00	-0.38	-0.23	0.00
Factor 3	-2.27	-0.05	0.00	-7.38	-0.08	0.00	0.04	0.03	0.03	0.08	0.07	0.00
Factor 4	7.35	0.11	0.00	*	*	*	*	*	*	*	*	*
R^2	0.0361			0.0492			0.0233			0.0337		
MODEL	23.06+ 3.90*FACT1+ 2.92*FACT2-2.27*FACT3+ 7.35*FACT4			245.99+ 25.32*FACT1+ 7.55*FACT2-7.38*FACT3			2.55+ 0.26*FACT1- 0.33*FACT2+ 0.04*FACT3			2.58+ 0.25*FACT1- 0.38*FACT2+ 0.08*FACT3		

* Variable rejected from the final model during previous iterations ($p > 0.05$)

Table 34

Multiple Regression Summary Table for New Factors (Science Only) – Showing Final Iterations

	SGP			CST			GPAS2			GPAS1		
	B	Beta	P	B	Beta	P	B	Beta	P	B	Beta	P
Factor 1	12.53	0.19	0.00	29.01	0.25	0.00	-0.10	-0.05	0.02	0.13	0.06	0.00
Factor 2	5.37	-0.10	0.00	10.56	-0.11	0.00	0.18	0.11	0.00	*	*	*
Factor 3	*	*	*	11.63	0.14	0.00	*	*	*	*	*	*
Factor 4	11.17	0.18	0.00	*	*	*	*	*	*	*	*	*
R^2	0.0874			0.0807			0.0078			0.0041		
MODEL	1.90+ 12.53*FACT1- 5.37*FACT2+ 11.17*FACT4			246.92+ 29.01*FACT1- 10.56*FACT2+ 11.63*FACT3			2.20-0.10*FACT1+ 0.18*FACT2			2.02+ 0.13*FACT1		

* Variable rejected from the final model during previous iterations ($p > 0.05$)

Research Question #6

Research question #6 asked, Are there indicators (or combinations of indicators) that can predict positive student outcomes? Stepwise-regression analysis was employed for research question #6 to determine what variables provide the greatest statistical contribution among the 29 indicators. The results to the analyses are reported subsequently in Tables 35-39.

For the data set as a whole, a model involving 20 indicators accounted for 6.35% (R^2 = 0.0635) of the variation in CST scores. For SGP, a model consisting of 17 indicators accounted for 4.67% (R^2 = 0.0467) of the variation. For GPAS1, a model consisting of 19 indicators accounted for 3.60% (R^2 = 0.0360) of the variation. In addition, GPAS2 had a model consisting of 20 indicators accounting for 2.93% (R^2 = 0.0293) of the variation.

Disaggregated by subject, stepwise-regression analysis yielded the largest R^2 in science, math, and history. Seventeen indicators accounted for 19.5% (R^2 = 0.195) of the variation in SGP in science, while 19 indicators accounted for 18.1% (R^2 = 0.181) of the variation in CST in math. A model with 17 indicators accounted for 18.1% (R^2 = 0.181) of the variation in CST for history. Furthermore, a model involving 19 indicators accounted for 16.7% (R^2 = 0.167) of the variation in CST for science.

Table 35

Step-wise Regression Summary Chart (All Subjects)

	R^2	# of indicators	Indicators
SGP	0.047	17	1.1.A, 1.2.B, 1.3.A, 1.4.A, 1.5.B, 2.1.A, 2.2.A, 2.2.B, 2.3.A, 2.3.B, 3.1.B, 3.1.C, 3.2.A, 3.3.A, 3.3.B, 3.3.D, 3.4.A.
CST	0.064	20	1.2.A, 1.2.B, 1.3.A, 1.5.B, 2.1.A, 2.2.B, 2.3.A, 2.3.B, 2.4.A, 3.1.A, 3.1.B, 3.1.C, 3.2.A, 3.3.A, 3.3.B, 3.3.D, 3.4.A, 3.4.C, 4.1.A, 4.1.B.
GPAS2	0.036	19	1.1.B, 1.3.A, 1.5.B, 2.2.A, 2.2.B, 2.3.A, 2.3.B, 2.4.A, 3.1.A, 3.1.B, 3.2.A, 3.2.B, 3.3.A, 3.3.B, 3.3.C, 3.4.A, 3.4.B, 3.4.C, 4.1.B.
GPAS1	0.029	20	1.1.A, 1.1.B, 1.3.A, 1.4.B, 1.5.A, 1.5.B, 2.2.A, 2.2.B, 2.3.A, 2.3.B, 3.1.B, 3.2.A, 3.2.B, 3.3.A, 3.3.B, 3.3.C, 3.3.D, 3.4.B, 3.4.C, 4.1.B.

Table 36

Step-wise Regression Summary Chart (ELA Only)

	R^2	# of indicators	Indicators
SGP	0.033	14	1.1.A, 1.1.B, 1.2.B, 1.3.A, 1.4.A, 1.5.A, 1.5.B, 2.1.A, 2.3.A, 3.1.C, 3.2.A, 3.3.B, 3.3.C, 4.1.A.
CST	0.075	20	1.1.A, 1.1.B, 1.2.A, 1.2.B, 1.4.A, 1.4.B, 1.5.A, 1.5.B, 2.1.A, 2.2.A, 2.2.B, 2.3.A, 2.3.B, 3.1.A, 3.1.B, 3.1.C, 3.3.B, 3.4.A, 3.4.B, 3.4.C.
GPAS2	0.094	16	1.1.B, 1.2.A, 1.4.A, 1.5.B, 2.1.A, 2.2.A, 2.2.B, 2.3.A, 2.4.A, 3.1.A, 3.2.A, 3.3.A, 3.3.B, 3.4.B, 3.4.C, 4.1.A.
GPAS1	0.083	19	1.1.B, 1.2.A, 1.2.B, 1.4.A, 1.5.B, 2.2.B, 2.3.A, 2.3.B, 2.4.A, 3.1.B, 3.1.C, 3.2.B, 3.3.A, 3.3.B, 3.3.C, 3.3.D, 3.4.A, 3.4.B, 3.4.C.

Table 37

Step-wise Regression Summary Chart (History Only)

	R^2	# of indicators	Indicators
SGP	0.105	13	1.1.B, 1.2.A, 1.2.B, 1.3.A, 1.4.B, 2.1.A, 2.2.A, 2.2.B, 2.3.A, 3.1.B, 3.1.C, 3.3.B, 4.1.A.
CST	0.181	17	1.1.B, 1.2.A, 1.3.A, 1.5.B, 2.1.A, 2.2.A, 2.3.A, 2.3.B, 3.1.A, 3.1.B, 3.1.C, 3.2.B, 3.3.B, 3.3.C, 3.3.D, 3.4.A, 3.4.C.
GPAS2	0.132	18	1.1.A, 1.2.A, 1.3.A, 1.4.A, 1.4.B, 1.5.B, 2.1.A, 2.2.B, 2.3.B, 2.4.A, 3.1.A, 3.1.B, 3.2.A, 3.2.B, 3.3.A, 3.3.B, 3.4.A, 3.4.C.
GPAS1	0.145	18	1.1.A, 1.1.B, 1.2.A, 1.2.B, 1.3.A, 1.4.A, 1.4.B, 2.1.A, 2.2.B, 2.3.A, 2.4.A, 3.1.B, 3.3.C, 3.3.D, 3.4.A, 3.4.C, 4.1.A, 4.1.B.

Table 38

Step-wise Regression Summary Chart (Math Only)

	R^2	# of indicators	Indicators
SGP	0.121	18	1.1.A, 1.2.A, 1.3.A, 1.4.A, 1.5.A, 1.5.B, 2.1.A, 2.2.A, 2.3.A, 2.3.B, 2.4.A, 3.1.B, 3.2.A, 3.2.B, 3.3.A, 3.3.D, 3.4.B, 4.1.B.
CST	0.181	19	1.2.B, 1.3.A, 1.5.A, 1.5.B, 2.1.A, 2.2.A, 2.3.A, 2.3.B, 2.4.A, 3.1.A, 3.1.B, 3.2.A, 3.2.B, 3.3.A, 3.3.B, 3.3.C, 3.4.A, 3.4.B, 4.1.A.
GPAS2	0.101	17	1.1.A, 1.3.A, 1.4.B, 1.5.A, 1.5.B, 3.1.A, 3.2.A, 3.2.B, 3.3.A, 3.3.B, 3.3.C, 3.3.D, 3.4.A, 3.4.B, 3.4.C, 4.1.A, 4.1.B.
GPAS1	0.100	13	1.4.A, 1.4.B, 1.5.A, 3.2.A, 3.2.B, 3.3.B, 3.3.C, 3.3.D, 3.4.A, 3.4.B, 3.4.C, 4.1.A, 4.1.B.

Table 39

Step-wise Regression Summary Chart (Science Only)

	R^2	# of indicators	Indicators
SGP	0.195	17	1.1.A, 1.1.B, 1.3.A, 1.4.A, 1.4.B, 1.5.A, 1.5.B, 2.1.A, 2.2.B, 2.4.A, 3.1.A, 3.1.B, 3.1.C, 3.3.B, 3.3.D, 3.4.A, 3.4.C.
CST	0.167	19	1.1.A, 1.1.B, 1.2.A, 1.2.B, 1.3.A, 1.5.B, 2.2.B, 2.3.A, 3.1.A, 3.1.B, 3.2.A, 3.2.B, 3.3.A, 3.3.B, 3.3.D, 3.4.A, 3.4.B, 3.4.C, 4.1.B.
GPAS2	0.114	15	1.2.B, 1.3.A, 1.4.A, 1.4.B, 1.5.A, 2.2.A, 2.2.B, 2.3.A, 2.4.A, 3.1.A, 3.2.A, 3.2.B, 3.3.B, 3.4.B, 4.1.B.
GPAS1	0.120	20	1.1.B, 1.2.B, 1.4.A, 1.4.B, 1.5.A, 1.5.B, 2.1.A, 2.2.A, 2.2.B, 2.3.A, 2.3.B, 2.4.A, 3.1.A, 3.1.C, 3.2.A, 3.2.B, 3.4.A, 3.4.B, 3.4.C, 4.1.A.

Chapter 5: Conclusions and Implications

This chapter will summarize the study through the lens of current research literature, draw conclusions and implications from the literature, and finally arrive at a series of recommendations for further research. The first section of this chapter will revisit the purpose of the study including the research questions posed as well as the research methodology. The following section will summarize the findings of the study in light of current research presented in Chapter 2. Conclusions from and implications of the results will then be presented in the next section. Finally, a number of recommendations for future research will be made.

Restatement of Purpose and Research Questions

The purpose of this study was to determine whether or not Green Dot's CRTF can account, in whole or in part, for differences in student outcomes. In so doing, the study sought to determine the validity of Green Dot's teacher effectiveness program as predictor of student achievement outcomes. This study was guided by the following research questions:

1. Are there differences in student outcomes (e.g., GPA and CST scores) based on a teacher's overall teacher effectiveness score?
2. Are there differences in student outcomes (e.g., GPA, CST, and SGP scores) based on a teacher's overall teacher observation score?
3. Are there differences in student outcomes (e.g., GPA, CST, and SGP scores) based on a teacher's score in a particular CRTF Domain?
4. Are there other observable constructs other than the five Domains identified by CRTF designers that can be arrived at through statistical analysis methods such as factor analysis?

5. Are there differences in student outcomes (e.g., GPA, CST, and SGP scores) based on a teacher's score in a particular CRTF construct arrived at through factor analysis?
6. Are there indicators (or combinations of indicators) that can predict positive student outcomes?

Summary of Key Findings

Study findings indicated that there were no remarkable relationships between a teacher's composite TES and student outcomes. Overall, SGP, CST, and TES yielded very weak positive relationships for all subjects. Taken separately for math and science, TES was moderately related to CST and SGP. Looking at history alone, TES was only moderately related to CST. In all cases, GPA for both semesters was not related to TES.

The findings for TObs were similar. TObs, a primary component of the composite TES, was weakly correlated with CST and SGP for all subjects. This was also true for the disaggregated data. As with TES, TObs was not related to GPA in either semester. Furthermore, no teacher effectiveness variable, TES or TObs, by itself accounted for more than nine percent of the variation in any of the student outcomes in the aggregated data (all subjects) and more than 16% of the variation in disaggregated data (separate subjects alone).

Green Dot's theoretical CRTF Domains also showed unremarkable relationships between student outcomes. Domains 1-4 by themselves were weakly related to all student outcomes. Consequently, no regression model involving the Domains accounted for variances that had any practical significance. There were no relationships among GPAS1, GPAS2, and any CRTF Domain. This was the case for both the aggregated and disaggregated data. Four additional CRTF Domains emerged through factor analysis. The new Factors were similar to the current CRTF Domains: factor 2 was identical to Domain 1 and factor 3 was identical to Domain 4.

These new Factors were all weakly related to SGP and CST. GPAS1, GPAS2, and the new Factors (1-4) were not related.

No combination of indicators was found to predict student achievement accounts. A number of regression models using combinations of CRTF indicators were determined using step-wise regression analysis; however, no combination accounted for a remarkable percentage of the variation in any variable. Upon disaggregation for science only, a 17-indicator model accounted for 19.5% of the variation in SGP. This represented the largest variation of any of the models.

Discussion of Results

No strong relationships. Research questions 1, 2, 3, and 5 sought to determine if there was a relationship between the CRTF, as a whole or in part, and student outcomes such as SGP, CST, and GPA. This study found essentially no remarkable relationships between the CRTF and student outcomes. The findings showed that the composite TES may be positively correlated, however, the relationships reported were extremely weak and inconsistent. Furthermore, the correlation values between TObs and student outcomes showed no correlations in both the aggregated and disaggregated data. This finding is largely consistent with research findings by Kimball et al. (2004), which found that a similar rubric based on the Danielson framework reported promising, yet inconsistent results between a teacher's effectiveness score in the framework and student outcomes. Furthermore, Goe's (2007) work on teacher practices found a number of positive correlations between teacher practice and student achievement; however, these relationships were not statistically or even practically significant. Furthermore, breaking up the CRTF into its smaller components, such as the pre-defined Domains created by the CRTF designers, also yielded no remarkable relationships between the rubric and student outcomes. New Factors arrived at through factor analysis were also not related to student outcomes.

Some moderate relationships seen in subject-specific data. In addition, only when the data were disaggregated for science and math alone did some moderate relationships emerge between student outcomes and the CRTF. This was not the case for the overall dataset when all subjects were combined. ELA, on the other hand, consistently showed no correlations when disaggregated from the entire data set. Some possible explanations for these differences include lack of alignment and pedagogical inconsistencies between teacher evaluation program and state standards (Heneman et al. 2006) or subjective and/or inconsistent ratings due to improperly calibrated evaluators (Wise et al., 1985), some of which may be the case at Green Dot. Furthermore, Clotfelter, et al (2006), studying the effect of teacher qualifications on student test scores, concluded that results are different depending on the subject matter.

Theoretical Domains are similar to Factors arrived at via factor analysis. Research question #4 sought to determine if additional Domains or Factors, other than the theoretical Domains originally defined by CRTF designers, could emerge through factor analysis. The factor analysis observed four Factors from the dataset that were similar to the theoretical Domains. Two of these Factors were identical to two theoretical Domains. This was an intriguing finding that seems to validate the underlying theoretical constructs created by the rubric designers (Danielson, 2007).

No predictable models discovered through regression analyses. Regression analysis was unable to determine a model that could account for variability that was practically significant from any student outcome metric. Clotfelter et. al. (2006) arrived at a similar conclusion when determining a regression model that accounted for variation in both reading and math scores based on two factors: teacher licensure test scores and teacher experience.

Implications of the Study

Darling-Hammond (2000), Hanushek (1992), and Wright et al. (1997), laid the foundation for research that regarded the teacher as one of the determining factors for student success and improvement. Therefore, it is vitally important to arrive at a valid tool that identifies effective teachers through proficiency in effective teaching practices. According to Goe (2007), however, there is no agreed upon definition of teacher effectiveness, let alone a standard measurement to determine what makes a teacher effective. This makes the development of teacher evaluation systems challenging due to the breadth of current ideas and evaluation rubrics available that have not shown significant relationships with student outcomes. The purpose of this study was to determine whether or not a particular rubric, Green Dot's CRTF, was related to three student outcomes (SGP, CST, and GPA). In so doing, the study would serve to validate the tool as a predictor for these student outcomes. The study also sought to define a predictive model of CRTF indicators and/or Domains that would accurately predict student outcomes such as SGP, CST, and GPA. A valid model that emerged from the study would provide additional rationale to develop professional development to bolster a teacher's proficiency in these variables.

The results of this study are consistent with the current body of literature surrounding student achievement outcomes and teacher effectiveness and suggest that teacher effectiveness, as measured by the CRTF, is not related to student outcomes (SGP, CST, and GPA). Furthermore, the results also suggest that both TES and TObs are not reliable predictors of student outcomes, as the study found no predictive regression models from Domains, Factors, or individual indicators in the CRTF that accounted for a practically significant amount of variation in any of the student outcomes.

Slightly higher correlations were reported when data were disaggregated by subject. This may imply a difference in professional development between subjects. CST disaggregated by history and science alone, reported moderate correlations with TES. A possible explanation for this inconsistency may be due to the non-educational factors inherent in standardized testing as described by Betebenner (2009) and Goe et al. (2008), such as socio-economic status, ethnicity, and language bias. SGP yielded the same differential gap between aggregated data and disaggregated data for math and science. No strong relationships emerged between SGP and TES or TObs. Higher correlations may imply effective coaching and support for one subject area over another. Furthermore, the quality of professional development in terms of alignment to tested standards, frequency, types of support, and other variables may contribute to these differences.

A student achievement measure that relates to teacher effectiveness appears to remain elusive. A major implication of this study is the possibility that student outcomes such as SGP, CST, and GPA are not adequate measures of teacher effectiveness. There is a gap in the research literature relating a teacher's effectiveness to his/her students' GPA. The results in this study show that TES and TObs based on the CRTF are not related to student GPA for either semester. CST scores may be valid for representing student proficiency at the end of the year as a snapshot, however, as pointed out by Betebenner (2011), CST scores may not be a valid measure of teacher effectiveness if used alone due to non-educational factors. Furthermore, for SGP, Briggs and Betebenner (2009) asserted that, unlike value-added measures, SGP scores were never intended to infer that growth is attributed to variations in school or even teacher quality. Despite this research however, Green Dot has shown internally that SGP has merit as it relates to effective teaching, reporting relationships between SGP and TES (K. Keelen, personal communication, July 3, 2014). For the purposes of determining teacher effectiveness, at least anecdotally, SGP remains an important component of the Teacher Effectiveness program.

The TObs may affect correlations between TES and student outcomes. The MET Project's (2013) work on composite measures involving different weightings for student survey, achievement gains, and observations found that a model where 50% of the teacher evaluation score was attributed to the teacher observation resulted in the lowest correlations with state tests and higher-order tests in ELA, yet provided the highest reliability among tested models. The findings in this study may align with this tradeoff, as Green Dot attributes 40% (tested teachers) and 55% (non-tested teachers) of the TES to classroom observations. Only weak to moderate correlations were reported from the dataset.

Teacher effectiveness, a relatively new body of literature, is scant with empirical research studies on specific evaluation rubrics that determine good teaching and student improvement (Kyriakides et al., 2006). This study contributes data consistent with other studies and adds Green Dot's CRTF to the current body of literature on multiple-measure teacher effectiveness rubrics and their relationship with student outcomes.

Limitations

The findings in this study are subject to at least three limitations. First, the secondary data explored represent only 1 year of teacher effectiveness data. The CRTF rubric language was constantly in flux at Green Dot as collective bargaining agreements between the district and the teacher's union proposed amendments and adjustments to the rubric language on a yearly basis. The decision was made to select only 1 years' worth of data due to possible inconsistent measurements between indicators over subsequent years. Additional years of data may have provided additional variability in the variables and/or provided data for longitudinal studies of the effects of the CRTF.

Second, not all CRTF indicators were individually ratable. As a result, the study was unable to incorporate certain indicators from Domain 4 or Domain 5 into the correlation and regression models. Consequently, the statistical analyses involving the relationships between theoretical Domains (research question #3), new Factor domains (research question #5) and student outcome metrics only accounted for the scores that administrators provided as the TObs. Quantitative student, peer, and administrator perception feedback surveys satisfied data from the indicators that were not included in the scored rubric. The scores for the indicators associated with perception surveys were aggregated as averages and weighted into the calculation of the TES based on teacher type. Because these indicators were not individually ratable, this presented a limitation in the full exploration of the rubric as entire Domains or indicators that were not part of the formal evaluation process could not be evaluated individually during the statistical analysis.

Third, it is conceivable that the secondary data had errors in the form of inter-rater reliability due to improper training or calibration. While this may not be the fault of the study's methodology due to the nature of secondary data, inter-rater reliability by school site may have affected the results of the data. Wise et al. (1985) described the issue of inconsistent scoring due to calibration issues of evaluators. At Green Dot, the TObs represents the largest factor (40% or greater) in determining TES and therefore would provide the largest contribution to the variability of the measure. Inconsistent scoring due to improperly calibrated evaluators may present issues with variability. Green Dot's certification protocol has been in flux over the course of the past several years, with the current system requiring administrators to certify once per year. Green Dot understands the importance of properly calibrating administrators as raters, but the process has been challenging to balance the need for inter-rater reliability and evaluator fatigue (K. Keelen, personal communication, July 3, 2014).

Recommendations for Future Research

To provide a more complete picture of Green Dot's CRTF and provide additional rationale for its continued development and improvement, it is recommended that further research be undertaken in the following areas.

The impact of different CRTF composites on correlation. Research should be conducted on the CRTF components (i.e., TObs, teacher's SGP, school's overall SGP, and stakeholder perception surveys) to determine an optimal model for weighting factors in the composite. Various composites should be compared and then tested to determine the model that best accounts for the variability in selected student outcomes. This was the work of the MET Project in 2013, where three different components were weighted differently among four separate models and then correlated with state test gains. Essentially, the study needs to be replicated using the components of the CRTF.

The effect of professional development on TES score. Moderate relationships among TES, TObs, and student outcomes emerged only when data were disaggregated for math, science, and, to a lesser extent, history. The nature of professional development offered between subject areas should be investigated further to determine if there is a difference in the quality, quantity, or content of the professional development being offered for various subject areas at Green Dot. Both quantitative and qualitative data in the form of surveys and/or focus groups may provide a possible explanation of the observed differences in moderate correlations for math and science versus no to weak correlations in ELA.

Teacher observation score and inter-rater reliability. The data showed that TObs was less correlated with student outcomes than the composite TES (both in aggregated and disaggregated data), which included TObs as well as other measures such as individual teacher SGP, overall school SGP, and stakeholder feedback. Since a large percentage of the TES is attributed to the TObs (40% for tested teachers and 55% for non-tested teachers), an exploratory research study involving only the TObs variable should be conducted. Furthermore, since administrators who have been trained by Green Dot are responsible for the TObs, research on inter-rater reliability should be conducted to determine the cause for variability within the variable and whether or not there is a calibration issue within the organization.

Summary

The concept of measuring teacher effectiveness continues to pose challenges. The purpose of this research was to determine if there was a relationship between student outcomes and teacher effectiveness, as defined by Green Dot's CRTF. In so doing, the study would serve to validate the use of the CRTF. While the research findings in this study may have been insignificant towards these ends, reporting no correlations between TES and student outcomes, the study provides a launching point for further research on Green Dot's rubric as well as similar rubrics. Tools such as the CRTF must continue to be refined based on research data. Quantitative studies such as the current study should be followed up by additional mixed-method studies to describe possible explanations for the differences in the data.

REFERENCES

American Federation of Teachers. (2013). *A guide for developing multiple measures for teacher development and evaluation*. Retrieved from https://www.aft.org/pdfs/teachers /devmultiplemeasures.pdf

Baker, B. D. (2011, September 2). Take your SGP and VAMit, damn it [Blog post]! Retrieved from http://schoolfinance101.wordpress.com/2011/09/02/take-your-sgp-and-vamit-damn-it/

Betebenner, D. (2009). Norm- and criterion-referenced student growth. *Educational Measurement: Issues and Practice*, *28*(4), 42-51. doi:10.1111/j.1745-3992.2009.00161.x

Betebenner, D. W. (2011). *A technical overview of the student growth percentile methodology: Student growth percentiles and percentile growth projections/trajectories*. Retrieved from http://www.nj.gov/education/njsmart/performance/SGP_Technical_Overview.pdf

Betts, J. R., Zau, A. C., & Rice, L. A. (2003). *Determinants of student achievement: New evidence from San Diego*. Retrieved from http://wwwww.ppic.org/content/pubs /report/R_803JBR.pdf

Borman, G. D., & Kimball, S. M. (2005). Teacher quality and education equality: Do teachers with higher standards-based evaluation ratings close student achievement gaps? *The Elementary School Journal*, *106*, 3-20. doi:10.1086/496904

Briggs, D., & Betebenner, D. (2009, April). Is growth in student achievement scale dependent? In J. Weeks, *Measuring and evaluating changes in student achievement: A conversation about technical and conceptual issues*. Symposium conducted at the National Council for Measurement in Education, San Diego, CA. Retrieved from http://dirwww.colorado.edu/education/faculty/derekbriggs/Docs/Briggs_Weeks_Is%20Growth%20in%20Student%20Achievement%20Scale%20Dependent.pdf

Camara, W., Kimmel, E., Scheuneman, J., & Sawtell, E. A. (2003). *Whose grades are inflated?* Retrieved from https://research.collegeboard.org/sites/default/files/publications /2012/7/researchreport-2003-4-whose-grades-are-inflated.pdf

Cavalluzzo, L. (2004). *Is National Board Certification an effective signal of teacher quality?* Retrieved from http://www.nbpts.org/sites/default/files/documents/research /Cavalluzzo_IsNBCAnEffectiveSignalofTeachingQuality.pdf

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2006). *Teacher-student matching and the assessment of teacher effectiveness (NBER Working Paper No. 11936*). Retrieved from http://www.nber.org/papers/w11936.pdf

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2007). *How and why do teacher credentials matter for student achievement?* Retrieved from http://www.nber.org/papers/w12828.pdf?new_window=1

Cohen, D. K., & Hill, H. C. (1998). *Instructional policy and classroom performance: The mathematics reform in California* (CPRE Research Report Series RR-39). Retrieved from http://www.cpre.org/images/stories/cpre_pdfs/rr39.pdf

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Wood, A. M., Weinfeld, F. D., & York, R. L. (1966). *Equality of educational opportunity* (ED 012 275). Washington, DC: Government Printing Office.

Cubberley, E. P. (1916). *Public school administration*. Retrieved from https://archive.org/details/publicschooladmi1922cubb

Danielson, C. (2007). *Enhancing professional practice: A framework for teaching* (2nd ed.). Princeton, NJ: The Danielson Group.

Danielson, C. (2011). *The framework for teaching evaluation instrument* (1st ed.). Princeton, NJ: The Danielson Group.

Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis Archives*, *8*(1), 1-44. doi: 10.3102/00346543073001089

Darling-Hammond, L. (2004). From “separate but equal” to “no child left behind:” The collision of new standards and old inequalities. In D. Meier, & G. Woods, *Many children left behind* (pp. 3-32). Boston: Beacon Press.

Darling-Hammond, L., Holtzman, D. J., Gatlin, S. J., & Heilig, J. V. (2005, October 12). Does teacher preparation matter? Evidence about teacher certification, Teach for America, and teacher effectiveness. *Education Policy Analysis Archives*, *13*(42). Retrieved from http://epaa.asu.edu/ojs/article/view/147/273

Dee, T. S. (2004). The race connection: Are teachers more effective with students who share their ethnicity? *Education Next, 2,* 52-59. Retrieved from http://educationnext.org/the-race-connection/

Derry, S. J. (1996). Cognitive schema theory in the constructivist debate. *Educational Psychologist*, *31*(3/4), 163-174. doi:10.1080/00461520.1996.9653264

Ehrenberg, R. G., Goldhaber, D. D., & Brewer, D. J. (1995). Do teachers’ race, gender, and ethnicity matter? Evidence from the National Educational Longitudinal Study of 1988. *Industrial and Labor Relations Review*, *48*, 547-561. doi:10.2307/2524781

Ferguson, R. F. (1991). Paying for public education: Evidence of how and why money matters. *Harvard Journal on Legislation*, *28*, 465-498.

Friedman, J. A. (2006). *An exploration of the role of teacher evaluation in a high-performing urban school: A case study* (Doctoral dissertation). Available from ProQuest Dissertations & Theses. (AAT No. 304971155)

Frome, P., Lasater, B., & Cooney, S. (2005). *Well-qualified teachers and high quality: Are they the same?* Retrieved from http://publications.sreb.org/2005/05V06_Research_Brief_high-quality_teaching.pdf

Gallagher, H. A. (2004). Vaughn elementary's innovative teacher evaluation system: Are teacher evaluation scores related to growth in student achievement? *Peabody Journal of Education*, *79*(4), 79-107. doi:10.1207/s15327930pje7904_5

Geiser, S., & Studley, R. (2002). UC and the SAT: Predictive validity and differential impact of the SAT I and SAT II at the University of California. *Educational Assessment*, *8*(1), 1-26. doi:10.1207/S15326977EA0801_01

Geisler, S., & Santelices, M. (2007). *Validity of high-school grades in predicting student success beyond the freshman year: High-school record vs. standardized tests as indicators of four-year college outcomes*. Retrieved from http://cshe.berkeley.edu/publications /validity-high-school-grades-predicting-student-success-beyond-freshman-yearhigh-school

George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference* (4th ed.). Boston, MA: Allyn & Bacon.

Gertsman, B. B. (n.d.). *StatPrimer* (ver. 6.4). Retrieved from http://www.sjsu.edu/faculty /gerstman/StatPrimer/

Glickman, C. (1985). *Supervision of instruction: A developmental approach*. Upper Saddle River, NJ: Pearson.

Goe, L. (2007). *The link between teacher quality and student outcomes: A research synthesis.* Retrieved from http://files.eric.ed.gov/fulltext/ED521219.pdf

Goe, L., Bell, C., & Little, O. (2008). *Approach to evaluating teacher effectiveness: A research synthesis*. Retrieved from http://www.gtlcenter.org/sites/default/files/docs/EvaluatingTeachEffectiveness.pdf

Goe, L., Biggers, K., & Croft, A. (2012). *Linking teacher evaluation to professional development: Focusing on improving teaching and learning*. Washington, DC: National Comprehensive Center for Teacher Quality.

Goldhaber, D., & Anthony, E. (2007). Can teacher quality be effectively assessed? National board certification as a signal of effective teaching. *The Review of Economics and Statistics*, *89*, 134-150. doi:10.1162/rest.89.1.134

Goldhammer, R. (1969). *Clinical supervision (special for the supervision of teachers)*. New York, NY: Holt, Rinehart, and Winston.

Hanushek, E. A. (1992). The trade-off between child quantity and quality. *The Journal of Political Economy*, *100*, 84-117. doi:10.1086/261808

Harris, D. N. (2011). *Value-added measures in education: What every educator needs to know*. Cambridge, MA: Harvard Education Press.

Heneman, H. G., Milanowski, A., Kimball, S. M., & Odden, A. (2006). *Standards-based teacher evaluation as a foundation for knowledge- and skill-based pay*. Retrieved from http://usny.nysed.gov/rttt/docs/policybrief.pdf

Holtzapple, E. (2003). Criterion-related validity evidence for a standards-based teacher evaluation system. *Journal of Personnel Evaluation in Education*, *17*(1), 207-219. doi:10.1007/s11092-005-2980-z

Howard, B. B., & McColskey, W. H. (2001). Evaluating experienced teachers. *Educational Leadership*, *58*(5), 48-51.

Hunter, M. (1982). *Mastery teaching*. El Segundo, CA: TIP Publications.

Kane, T. J., Rockoff, J. E., & Staiger, D. O. (2007). What does certification tell us about teacher effectiveness? Evidence from New York City. *Economics of Education Review*, *27*, 615-631. doi:10.1016/j.econedurev.2007.05.005

Kimball, S. M., White, B., Milanowski, A. T., & Borman, G. (2004). Examining the relationship between teacher evaluation and student assessment results in Washoe County. *Peabody Journal of Education*, *79*(4), 54-78. doi:10.1207/s15327930pje7904_4

Kyriakides, L., Demetriou, D., & Charalambous, C. (2006). Generating criteria for evaluating teachers through teacher effectiveness research. *Educational Research*, *48*(1), 1-20. doi:10.1080/00131880500498297

Leana, C. R., & Pil, F. K. (2006). Social capital and organizational performance: Evidence from urban public schools. *Organization Science*, *17*, 353-366. doi:10.1287/orsc.1060.0191

MET Project. (2012). *Asking students about teaching: Student perception surveys and their implementation*. Retrieved from http://www.metproject.org/downloads /Asking_Students_Practitioner_Brief.pdf

MET Project. (2013). *Ensuring fair and reliable measures of effective teaching*. Retrieved from http://www.metproject.org/downloads/MET_Ensuring_Fair_and_Reliable_Measures_Practitioner_Brief.pdf

Mabry, L., & Margolis, J. (2006, September 14). NCLB: Local implementation and impact in southwest Washington State. *Educational Policy Analysis Archives*, *14*(23), 1-38. doi: 10.14507/epaa.v14n23.2006

Marzano, R. J., Frontier, T., & Livingston, D. (2011). *Effective supervision: Supporting the art and science of teaching*. Alexandria, VA: ASCD.

McCaffrey, D. F., Lockewood, J. R., Koretz, D. M., & Hamilton, L. S. (2003). *Evaluating value-added models for teacher accountability*. Santa Monica, CA: RAND Corporation.

Monk, D. H. (1994). Subject area preparation of secondary mathematics and science teachers and student achievement. *Economics of Education Review*, *13*, 125-145. doi:10.1016/0272-7757(94)90003-5

Multiple choices: Options for measuring teaching effectiveness. (2013). Retrieved from http://www.rand.org/education/projects/measuring-teacher-effectiveness/multiple-choices.html

No Child Left Behind Act of 2001, 20 U.S.C. § 7801 *et seq.* (2002).

Noble, J., & Sawyer, R. (2002). *Predicting different levels of academic success in college using high school GPA and ACT composite score*. Retrieved from http://www.act.org /research/researchers/reports/pdf/ACT_RR2002-4.pdf

Odden, A., & Kelly, J. (2008). *Strategic management of human capital in public education*. Retrieved from http://www.aypf.org/documents/StrategicManagementHC.pdf

Paley, A. R. (2007, March 14). "No child" target is out of reach. *The Washington Post*. Retrieved from http://www.washingtonpost.com/wp-dyn/content/article/2007/03/13 /AR2007031301781.html

Phelps, R. P. (2011). *The effect of testing on achievement: Meta-analyses and research summary, 1910-2010*. Retrieved from http://www.nonpartisaneducation.org/Review/Resources /QuantitativeList.pdf

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Economerica*, *73*, 417-458. Retrieved from http://www.econ.ucsb.edu/~jon/Econ230C/HanushekRivkin.pdf

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review*, *94*, 247-252. doi:10.1257/0002828041302244

Sanchez, E. (2013). *Differential effects of using ACT College Readiness Assessment Scores and high school GPA to predict first year college GPA among racial, ethnic, gender, and income groups*. Iowa City, IA: ACT Research Report Series.

Sanders, W. L., & Horn, S. P. (1998). Research findings from the Tennessee Value-Added Assessment System (TVAAS) database: Implications for educational evaluation and research. *Journal of Personnel Evaluation in Education*, *12*(3), 247-256. doi:10.1023 /A:1008067210518

Sanders, W. L., & Rivers, J. C. (1996). *Cumulative and residual effects of teachers on future student academic achievement*. Retrieved from http://heartland.org/sites/all/modules /custom/heartland_migration/files/pdfs/3048.pdf

Sawyer, R. (2013). Beyond correlations: Usefulness of high school average and ACT scores in making college admissions decisions. *Applied Measurement in Education*, *26*(2), 89-112. doi:10.1080/08957347.2013.765433

Slaughter, R. (2008, September). *Measuring middle school achievement growth with student growth percentile methodology*. Paper presented at the Conference of the Northeastern Educational Research Association, Storrs, CT. Retrieved from http://digitalcommons.uconn.edu/nera_2008/20

Stallings, J., Robbins, P., Presbrey, L., & Scott, J. (1986). Effects of instruction based on the Madeline Hunter model on students' achievement: Findings from a follow-through project. *The Elementary School Journal*, *86*(5), 571-587. doi:10.1086/461468

Strauss, V. (2011, April 15). Myths of standardized testing. *Washington Post*.

Taylor, F. W. (1911). *The principles of scientific management*. New York, NY: Harper & Brothers.

Taylor, R. (1990). Interpretation of the correlation coefficient: A basic review. *Journal of Diagnostic Medical Sonography*, *6*(1), 35-39. doi:10.1177/875647939000600106

Tracy, S. J. (1995). How historical concepts of supervision relate to supervisory practices today. *Clearing House*, *68*(5), 320325.

Tryfos, P. (1998). *Methods for business analysis and forecasting: Text & cases*. Retrieved from http://www.yorku.ca/ptryfos/f1400.pdf

Tucker, P. D., & Stronge, J. H. (2005). *Linking teacher evaluation and student learning*. Alexandria, VA: Association for Supervision and Curriculum Development.

U.S. Department of Education. (2004). *Testing: Frequently asked questions*. Retrieved from http://www2.ed.gov/nclb/accountability/ayp/testing-faq.html

U.S. Department of Education. (2009). *Race to the top executive summary* (RTTT). Washington, DC: Government Printing Office.

U.S. Department of Education. (2010). *Blueprint for reform: The reauthorization of the elementary and secondary education act*. Washington, DC: Government Printing Office.

Vygotsky, L. S. (1986). *Thought and language*. Cambridge, MA: The MIT Press.

Weisberg, D., Sexton, S., Mulhern, J., Keeling, D., Schunck, J., Palcisco, A., & Morgan, K. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness*. Retrieved from http://widgeteffect.org/downloads /TheWidgetEffect.pdf

Wenglinsky, H. (2000). *How teaching matters: Bringing the classroom back into discussions of teacher quality*. Retrieved from http://www.ets.org/Media/Research/pdf /PICTEAMAT.pdf

Wise, A. E., Darling-Hammond, L., & Bernstein, H. T. (1985). Teacher evaluation: A study of effective practices. *The Elementary School Journal*, *86*(1), 60-121. doi:10.1086/461437

Wright, S. P., Horn, S. P., & Sanders, W. L. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education*, *11*, 57-67. doi:10.1023/A:1007999204543

APPENDIX A

IRB Approval

PEPPERDINE UNIVERSITY

Graduate & Professional Schools Institutional Review Board

May 30, 2014

Narciso Aguda

Protocol #: N0514D01
Project Title: Is the College Ready Teaching Framework Related to Student Achievement

Dear Mr. Aguda,

Thank you for submitting the Non-Human Subjects Verification Form and supporting documents for your above referenced project. As required by the Code of Federal Regulations for the Protect for Human Subjects (Title 45 Part 46) any activity that is research and involves human subjects requires review by the Graduate and Professional Schools IRB (GPS-IRB).

After review of the Non-Human Subjects Verification Form and supporting documents, GPS IRB has determined that your proposed research[1] activity does not involve human subjects. Human subject is defined as a living individual about whom an investigator (whether professional or student) conducting research obtains (1) data through intervention or interaction with the individual, or (2) identifiable private information. (45 CFR 46102(f))

As you are not obtaining either data through intervention or interaction with living individuals, or identifiable private information, then the research activity does not involve human subjects, therefore GPS IRB review and approval is not required of your above reference research.

We wish you success on your non-human subject research.

Sincerely,

Dr. Thema Bryant-Davis
Chair, Graduate and Professional Schools IRB
Pepperdine University

cc: Dr. Lee Kats, Vice Provost for Research and Strategic Initiatives
Mr. Brett Leech, Compliance Attorney
Dr. Farzin Madjidi, Faculty Chair

[1] *Research m*eans a systematic investigation, including research development, testing and evaluation, designed to develop or contribute to generalizable knowledge. Activities which meet this definition constitute research for purposes of this policy, whether or not they are conducted or supported un-der a program which is considered research for other purposes. (45 CFR 46.102(d)).

APPENDIX B

The College-Ready Teaching Framework

Domain 1: *Data-Driven Planning and Assessing Student Learning*

Standard	Indicators	Does Not Meet Standard Level I	Partially Meets Standard Level II	Meets Standard Level III	Exemplifies Standard Level IV	Indicator At-A-Glance
1.1 Establish standards-based learning objectives and assessments	A) Selection of learning objectives	Learning objective(s) are missing a specific level of cognition or content. AND Learning objective(s) are misaligned (do not progress toward mastery of content standards).	Learning objective(s) are missing either a specific level of cognition or content. OR Learning objective(s) are misaligned (do not progress toward mastery of content standards).	Learning objective(s) include both specific levels of cognition and content. AND Learning objective(s) are aligned to and progress toward mastery of content standards.	*All of level 3 and...* Learning objective(s) exceed level of cognition or increase level of challenge required by content standards.	• *Do the objective(s) contain level of cognition and content?* • *Are the objective(s) at an appropriate level of rigor and scaffolds toward success on standard(s)?* *NOTE: Proving behavior is no longer assessed in this indicator. Now, it is assessed in 1.1B.*
	B) Measurability of learning objectives	Proving behavior does not measure the learning objective(s).	Proving behavior measures the learning objective(s). AND Proving behavior uses only general criteria for measuring success.	Proving behavior measures the learning objective(s). AND Proving behavior includes specific criteria (quantitative or qualitative) for measuring success.	*All of level 3 and...* Proving behavior is measured by multiple methods.	• *Is the proving behavior aligned with the learning objective(s)?* • *How is success on the proving behavior measured?*
1.2 Organize instructional plans to promote standards-based, cognitively engaging learning for students	A) Designing and sequencing of learning experiences	The design of the learning experiences is not aligned to the learning objective(s). AND Learning experiences are not sequenced to enable students to demonstrate independent mastery of the learning objective(s) through the gradual release of responsibility.	The design of the learning experiences is not aligned to the learning objective(s). OR Learning experiences are not sequenced to enable students to demonstrate independent mastery of the learning objective(s) through the gradual release of responsibility.	The design of the learning experiences is sequenced to enable students to demonstrate independent mastery of the learning objective(s) through the gradual release of responsibility.	*All of level 3 and...* The design of the learning experiences is differentiated to meet the needs of subgroups of students.	• *Are the learning experiences aligned to the learning objective(s)?* • *Are the learning experiences sequenced using gradual release of responsibility?* • *Level IV: Are learning experiences differentiated?* *NOTE: Instructional pacing is not assessed in this indicator. It is assessed using the phrase 'appropriate time' in indicator 1.2B*
	B) Creating cognitively engaging learning experiences for students	Instructional plans do not provide opportunity for cognitively engaging learning experiences throughout the lesson cycle.	Instructional plans include cognitively engaging learning experiences but without appropriate time and support throughout the lesson cycle.	Instructional plans include cognitively engaging learning experiences throughout the lesson cycle, and each learning experience provides appropriate time and support.	*All of level 3 and...* Instructional plans provide differentiated, cognitively engaging learning experiences for subgroups of students.	• *Are learning experiences consistently cognitively engaging?* • *Does the teacher plan appropriate time and support for students to fully engage in each learning experience?* • *Level IV: Are learning experiences differentiated?*
1.3 Use student data to guide planning	A) Lesson design guided by data	The teacher does not use student data to guide or inform planning.	The teacher uses student data to inform planning of content organization ***or*** instructional strategies. OR The teacher uses student data to inform planning that meets the needs of the whole class.	The teacher uses student data to inform planning of content organization ***and*** instructional strategies. AND The teacher uses student data to inform planning that meets the needs of subgroups of students.	*All of level 3 and...* The teacher cites instructional strategies to meet the needs of individual students.	• *Does the teacher use data to inform content organization and instructional strategies?* • *Is the data used to inform planning for the whole class, subgroups, or individual students?*

Domain 1: *Data-Driven Planning and Assessing Student Learning*

Standard	Indicators	Does Not Meet Standard Level I	Partially Meets Standard Level II	Meets Standard Level III	Exemplifies Standard Level IV	Indicator At-A-Glance
1.4 Use knowledge of subject matter content/skills and learning processes to plan for student learning	A) Knowledge of subject matter to identify pre-requisite knowledge	The teacher does not accurately identify or address the prerequisite knowledge and skills to achieve the standard/learning objective(s). OR The teacher does not include opportunities to activate prerequisite knowledge. OR The teacher does not include strategies to address potential gaps for whole group of students.	The teacher accurately identifies the prerequisite knowledge and skills to achieve the standard/learning objective(s). AND The teacher includes opportunities to activate prerequisite knowledge. AND The teacher includes strategies to address potential gaps for ***whole groups*** of students.	The teacher accurately identifies the prerequisite knowledge and skills to achieve the standard/learning objective(s). AND The teacher includes opportunities to activate prerequisite knowledge. AND The teacher includes strategies to address potential gaps for ***subgroups*** of students.	*All of level 3 and…* The teacher uses knowledge to address potential gaps for individual students.	• *Does the teacher identify pre-requisite knowledge and skills?* • *Does the teacher plan opportunities to activate previous knowledge?* • *Does the teacher plan strategies for the whole class, subgroups, or for individual students?* *NOTES: Activation of prior knowledge, which was previously not assessed, is now assessed in this indicator.*
	B) Addresses common content misconceptions	The teacher does not anticipate common student misconceptions and does not include strategies to ensure students recognize and address these misconceptions to master the standard/learning objective(s).	The teacher anticipates common student misconceptions but does not include strategies to ensure students recognize and address these misconceptions to master the standard/learning objective(s).	The teacher anticipates common student misconceptions and includes strategies that ensure students recognize and address these misconceptions to master the standard/learning objective(s).	*All of level 3 and…* The teacher includes opportunities for students to uncover and correct their own misconceptions.	• *Does the teacher anticipate common student misconceptions?* • *Does the teacher plan strategies to address student misconceptions?* • *Level IV: Do students uncover and correct their own misconceptions?* *NOTE: The language of the indicator ("standard/learning objective") allows teachers to address misconceptions in this OR future lessons.*
1.5 Design assessments to ensure student mastery	A) Selection and progression of assessments	Formative assessments are not aligned to the learning objective(s). OR Formative assessments are not planned.	The formative assessments are inconsistently aligned to the learning objective(s). OR Formative assessments do not yield actionable data. OR Formative assessments are planned for a single component of the lesson cycle.	A variety of formative assessments are selected to yield actionable data about progress towards mastery of the learning objective(s). AND Formative assessments are planned for different components of the lesson cycle, progressing towards student mastery of the learning objective(s).	*All of level 3 and…* The formative assessments are differentiated to yield actionable data about subgroups of students.	• *Are formative assessments aligned to the learning objective(s)?* • *Are formative assessments planned throughout the lesson?* • *Are a variety of formative assessment techniques used?* • *Do formative assessments yield actionable data?* • *Level IV: Is actionable data provided about the whole class or for subgroups?* *NOTE: Now, this indicator only assesses formative checks for understanding. The proving behavior is assessed in 1.1B.*
	B) Planned response to assessment data	The teacher has not planned a response to data from formative assessments.	The teacher inconsistently plans responses to data from formative assessments.	The teacher plans to adjust instruction based on the data from each formative assessment.	*All of level 3 and…* The teacher provides opportunities for students to use formative assessments to reflect on current progress toward the learning objective(s) or to determine next steps to extend learning.	• *Does the teacher plan to adjust instruction for each formative assessment?* • *Level IV: Do students reflect on their own progress and determine next steps to extend learning?* *NOTE: If not included explicitly in the lesson plan, this indicator may be assessed via verbal questioning in the observation pre-conference.*

Domain 2: *The Classroom Learning Environment*

Standard	Indicators	Does Not Meet Standard Level I	Partially Meets Standard Level II	Meets Standard Level III	Exemplifies Standard Level IV	Indicator At-A-Glance
2.1 Creates a classroom/ community culture of learning	A) Value of effort and challenge	The teacher's words and actions provide little or no encouragement for academic learning or convey low expectations for student effort. Students do not consistently persist in completing assigned work.	The teacher's words and actions emphasize compliance and completion of work. Students seek to complete tasks without consistent focus on learning or persistence toward quality work.	The teacher's words and actions promote belief in student ability and high expectations for student effort. Students consistently expend effort to learn and persist in producing high quality work.	*All of level 3 and...* Students assume responsibility or take initiative for producing high quality work, holding themselves, and each other, to high standards of performance.	• *Does the teacher emphasize completion and compliance, or learning and quality work?* • *Do students focus on completing assignments or producing quality work?* • *Level IV: Do students assume responsibility and take initiative for high quality work?*
2.2 Manage student behavior through clear expectations and a balance of positive reinforcement, feedback, and redirection	A) Behavioral expectations	It is evident that the teacher did not teach standards for student behavior. OR Student behavior does not contribute to an academic environment.	The teacher inconsistently communicates standards for student behavior. OR Student behavior inconsistently contributes to an academic environment.	The teacher consistently communicates clear, high standards for student behavior. AND Student behavior contributes to an academic environment.	The teacher has established clear, high standards for student behavior. Without being prompted, students articulate or promote behavioral expectations that support the classroom's academic environment.	• *Does the teacher communicate clear and high standards for student behavior?* • *Does student behavior contribute to an academic environment?* • *Level IV: Do students articulate and promote behavioral expectations without prompting?*
	B) Response to behavior	The teacher does not respond to misbehavior when necessary, or the response is repressive or disrespectful of student dignity.	The teacher's verbal or non-verbal response to student behavior is inconsistent. OR Teacher's verbal or non-verbal response is focused on the whole-class. OR Teacher emphasizes consequences over positive reinforcement.	The teacher's verbal or non-verbal response to student behavior is consistent, respectful, proactive, and includes redirection, feedback or positive reinforcement to specific students.	*All of level 3 and...* Students appropriately respond to or redirect each other's behavior.	• *Does the teacher consistently respond to both positive and negative student behavior?* • *Does the teacher respond to the whole class or to specific students?* • *Is the teacher respectful of student dignity?* • *Level IV: Do students respond to and redirect each other?*
2.3 Establish a culture of respect and rapport which supports students' emotional safety	A) Interactions between teacher and students	The teacher's interactions with some students are negative, demeaning, or inappropriate to the age and needs of the students in the class. OR Students exhibit disrespect for the teacher.	The teacher's interactions with students inconsistently demonstrate respect and positivity, or are not consistently appropriate for the age and needs of students in the class. OR Students inconsistently exhibit respect for the teacher.	The teacher's interactions with students are respectful, positive, and appropriate for the age and needs of the students in the class. AND Students exhibit respect for the teacher.	*All of level 3 and...* The teacher's interactions demonstrate a positive rapport with individual students.	• *Does the teacher interact with students in a respectful, positive and appropriate manner?* • *Do students exhibit respect for the teacher?* • *Level IV: Is there evidence of rapport between the teacher and individual students?*
	B) Student interactions with each other	Student interactions are impolite and disrespectful, which interferes with learning for some students.	Student interactions are generally polite and respectful, but students do not support each other's learning.	Student interactions are polite and respectful, and students support each other's learning.	*All of level 3 and...* Students encourage each other individually.	• *Do students interact polite and respectfully with each other?* • *Do students support each other's learning?* • *Level IV: Do students encourage each other?*
2.4 Use smooth and efficient transitions, routines, and procedures	A) Routines, procedures, and transitions	The teacher has not established or does not enforce routines, procedures, and transitions, resulting in a loss of instructional time.	The teacher has established some routines, procedures, and transitions; however, some may be missing or inconsistently enforced, resulting in the loss of instructional time.	The teacher has established and enforces routines, procedures, and transitions that maximize instructional time.	*All of level 3 and...* With minimal prompting, students effectively facilitate some routines, procedures, and transitions.	• *Has the teacher established routines, procedures, and transitions?* • *Do routines, procedures, and transitions maximize instructional time?* • *Level IV: Do students facilitate any routines, procedures or transitions?*

Domain 3: *Instruction*

Standard	Indicators	Does Not Meet Standard Level I	Partially Meets Standard Level II	Meets Standard Level III	Exemplifies Standard Level IV	Indicator At-A-Glance
3.1 Communicate learning objectives to students	A) Communication of the learning objectives of the lesson	The teacher does not explain the learning objective(s).	The teacher initially explains the learning objective(s) but does not refer to the objective(s) throughout the lesson. OR Students cannot articulate what they are expected to learn.	The teacher explains the learning objective(s) and refers back to it throughout the lesson. AND Students are able to articulate what they are expected to learn.	*All of level 3 and...* Students are able to articulate the relevance of the learning objective(s).	• *Does the teacher explain the learning objective?* • *Does the teacher refer back to learning objective throughout the lesson?* • *Level IV: Can students articulate the learning objective and its relevance?*
	B) Connections to prior and future learning experiences	The teacher makes limited connections between current learning objective(s) and the students' prior or future learning.	The teacher makes connections between the current learning objective(s) and the students' prior **or** future learning. Connections are vague or based on connections to assessments and grades.	The teacher makes connections between the current learning objective(s) and the students' prior **and** future learning to further student understanding of the content material within or outside of the discipline or unit.	The teacher facilitates as students build connections between the current learning objective(s) and their prior and future learning. Students make explicit connections within or outside of the discipline or unit.	• *Does the teacher connect the learning objective prior/future learning?* • *Are connections based on assessments/grades or on content within or outside the unit?* • *Level IV: Does the teacher facilitate while students make connections?*
	C) Criteria for success	The teacher does not establish criteria for successfully demonstrating attainment of the learning objective(s).	The teacher mentions but does not clearly explain the criteria for successfully demonstrating attainment of the learning objective(s).	The teacher clearly articulates the criteria for successfully demonstrating attainment of the lesson objective(s). AND Students are able to articulate the criteria for successfully demonstrating attainment of the learning objective(s).	*All of level 3 and...* The teacher solicits student ideas to define or affirm the criteria for successfully demonstrating attainment of the learning objective(s).	• *Does the teacher clearly articulate success criteria?* • *Can students articulate success criteria?* • *Are students involved in defining or affirming the success criteria?* *NOTE: Now, this indicator only focuses on the success criteria for the proving behavior, as opposed to success criteria for every activity.*
3.2 Facilitates Instructional Cycle	A) Executes lesson cycle	The teacher executes a lesson cycle that is inappropriately paced. AND The teacher does not execute a lesson cycle that gradually releases responsibility.	The teacher executes a lesson cycle that is inappropriately paced. OR The teacher does not execute a lesson cycle that gradually releases responsibility.	The teacher executes an appropriately paced lesson cycle that gradually releases responsibility so that students can independently master the learning objective(s).	*All of level 3 and...* To address the learning needs of subgroups, the teacher adapts the pacing or the release of responsibility.	• *Does the teacher appropriately pace the lesson?* • *Does the lesson gradually release responsibility to the students?* • *Level IV: Does the teacher adapt the pacing or release of responsibility for subgroups?*
	B) Cognitive Level of Student Learning Experiences	Learning experiences are not cognitively engaging. OR Learning experiences do not match the level of rigor required to attain mastery of the learning objective(s).	Some learning experiences are cognitively engaging. OR Some learning experiences match the level of rigor required to attain mastery of the learning objective(s).	Learning experiences throughout the lesson cycle are cognitively engaging. AND Learning experiences consistently match the level of rigor required to attain mastery of the learning objective(s).	*All of level 3 and...* Learning experiences require student thinking that exceeds the level of cognition or increases the level of challenge required by content standards.	• *Are learning experiences consistently cognitively engaging?* • *Is the rigor of the learning experiences aligned to the learning objective?* • *Level IV: Does student thinking exceed the level of cognition of cognitive challenge required by the standards?*

Domain 3: *Instruction*

Standard	Indicators	Does Not Meet Standard Level I	Partially Meets Standard Level II	Meets Standard Level III	Exemplifies Standard Level IV	Indicator At-A-Glance
3.3 Implementation of instructional strategies	A) Questioning	Many questions posed by the teacher do not move student thinking toward mastery of the learning objective(s). OR Most of the questions posed by the teacher require little cognitive challenge.	The teacher poses questions to a small number of students in the class. OR The teacher inconsistently scaffolds questions toward cognitive challenge and mastery of the learning objective(s).	The teacher poses questions to a wide range of students that are scaffolded toward cognitive challenge and mastery of the learning objective(s). AND The teacher uses strategies to enable students to correctly answer questions and extend or justify their thinking.	*All of Level 3 and…* Students pose questions that require cognitive challenge. OR Students initiate questions to further other students' understanding of the content.	• *Are questions aligned to the learning objective?* • *Are questions scaffolded toward cognitive challenge?* • *Are questions posed to a few students or to a wide range of students?* • *Are students required to correctly answer questions?* • *Are students required to extend or justify their thinking?* • *Level IV: Are students posing cognitively challenging questions?*
	B) Academic Discourse	The teacher does not require students to use the language of the discipline, discuss academic ideas, or justify their reasoning. OR The teacher provides minimal opportunities for student discussion.	The teacher inconsistently requires students in whole class or small group conversations to use the language of the discipline, discuss academic ideas, **or** justify their reasoning. OR Academic discourse is limited to a small number of students.	The teacher facilitates conversations in whole class and small group settings that require all students to consistently use the language of the discipline, discuss academic ideas, **and** justify their reasoning.	Students facilitate whole class or small group discussions and consistently use the language of the discipline, discuss academic ideas, and justify their reasoning.	• *How many students participate in academic discourse?* • *Do students participate in academic discourse consistently (in all activities and in all settings)?* • *Level IV: Do students or the teacher facilitate whole class or small group discussion?*
	C) Group structures	The structure and size of grouping arrangements do not move students toward mastery of the learning objective(s).	The structure and size of grouping arrangements inconsistently move students toward mastery of the learning objective(s). OR Students inconsistently participate within all group structures.	The structure and size of grouping arrangements move students toward mastery of the learning objective(s). AND Students actively participate within all group structures.	*All of level 3 and…* The teacher differentiates grouping arrangements in order to maximize learning for individual students. Students rely on each other to work through challenging activities and hold themselves and each other accountable for individual or group work.	• *Do the structure and size of grouping arrangements facilitate students' mastery of the learning objective?* • *Do students consistently and actively participate within the group structures?* • *Level IV: Is grouping differentiated?* • *Level IV: Do students rely on each other to work through challenging activities, holding each other accountable?*

	D) Resources and instructional materials	Resources and instructional materials are unsuitable to the lesson objective(s), distract from or interfere with student learning, or do not promote cognitive engagement.	Resources and instructional materials are partially suitable to the lesson objective(s). Resources and materials only partially promote cognitive engagement.	Resources and instructional materials are suitable to the lesson objective(s), support attainment of the learning objective(s), and promote cognitive engagement.	*All of level 3 and…*
					Resources and instructional materials require cognitive engagement. Students choose, adapt, or create materials to extend learning.
						• *Do resources and instructional materials:*
						  ○ *Suite the learning objective?*
						  ○ *Support attainment of the learning objective?*
						  ○ *Promote or require cognitive engagement?*
						• *Level IV: Do students choose, adapt, or create materials to extend learning?*

Domain 3: *Instruction*

Standard	Indicators	Does Not Meet Standard Level I	Partially Meets Standard Level II	Meets Standard Level III	Exemplifies Standard Level IV	Indicator At-A-Glance
3.4 Monitoring student learning during instruction	A) Checking for understanding and adjusting instruction	The teacher does not check for students' understanding of the learning objectives during the lesson. OR The teacher does not adjust instruction based on the data.	The teacher inconsistently checks for understanding throughout the lesson cycle. OR The checks do not yield actionable data on students' progress toward the learning objective(s). OR The teacher inconsistently or ineffectively adjusts instruction based on the data.	The teacher checks for understanding using varied techniques throughout the lesson cycle to yield actionable data on students' progress toward the learning objective(s). AND The teacher adjusts instruction based on the data to meet students' learning needs.	*All of level 3 and...* The teacher implements differentiated instruction and continued checks for understanding based on the progress of subgroups toward mastery of the learning objective(s).	• *Does the teacher check for understanding throughout the lesson cycle?* • *Does the teacher use varied techniques to check for understanding?* • *Do checks for understanding yield actionable data?* • *Does the teacher effectively adjust instruction based on data from checks for understanding?* • *Level IV: Does the teacher implement differentiated instruction for subgroups based on data from checks for understanding?* *NOTE: This indicator is aligned to 1.5A.*
	B) Feedback to students	The teacher does not provide feedback to students. OR Feedback does not advance students toward mastery of the learning objective(s).	The teacher provides feedback but not throughout the lesson cycle. OR Feedback inconsistently advances students toward attainment of the learning objective(s).	The teacher provides feedback throughout the lesson cycle that is specific and timely. AND Feedback consistently advances students toward attainment of the learning objective(s).	*All of level 3 and...* Students provide specific feedback to one another.	• *Does the teacher provide feedback throughout the lesson cycle?* • *Is feedback specific and timely?* • *Does feedback advance students toward the learning objective?* • *Level IV: Do students provide specific feedback to one another?*
	C) Self- monitoring	The teacher does not provide students with opportunities to engage in self- monitoring of their own progress or thinking.	The teacher provides students with limited opportunities for self-monitoring exercises.	The teacher provides students with opportunities for self-monitoring exercises that move students towards a deeper mastery of the objective(s).	Students self-monitor without the direction of the teacher. AND Students judge their own performance relative to success criteria.	• *Does the teacher provide opportunities for self-monitoring?* • *Does self-monitoring move students toward mastery of the learning objective?* • *Level IV: Do students self-monitor and judge their own performance?* *NOTE: Goal setting has been removed from this indicator.*

Domain 4: *Developing Professional Practice*

Standard	Indicators	Does Not Meet Standard Level I	Partially Meets Standard Level II	Meets Standard Level III	Exemplifies Standard Level IV
4.1 Engage in critical reflection, constantly revising practice to increase effectiveness	A) Accuracy	The teacher does not know the degree to which a lesson was effective or achieved its instructional goals, or profoundly misjudges the success of a lesson.	The teacher has a generally accurate impression of a lesson's effectiveness and success in meeting the instructional goals.	The teacher makes an accurate assessment of a lesson's effectiveness and success in meeting the instructional goals, citing general data to support the judgment.	*All of level 3 and...* The teacher cites specific data, and weighs the relative strengths of each data source. .
	B) Use in future planning	The teacher has limited suggestions for how the lesson could be improved.	The teacher makes general suggestions about how the lesson could be improved.	The teacher makes specific suggestions about how the specific lesson can be improved and general suggestions for improving the teaching practice as a whole.	*All of level 3 and...* The teacher predicts how the improvements will advance student learning in future lessons.
	C) Acceptance of feedback	The teacher is resistant to feedback from supervisors or colleagues and/or does not use the feedback to improve practice.	The teacher accepts feedback from supervisors and colleagues but may/may not use the feedback to improve practice.	The teacher welcomes feedback from supervisors and colleagues and uses the feedback to improve practice.	*All of level 3 and...* The teacher proactively seeks feedback on what has been implemented.
4.2 Engage in collaborative relationships with peers to learn and share best practices and ensure continuity in student learning	A) Participation in a professional community	The teacher avoids participating in the professional community activities or has strained relationships with colleagues that negatively impact the learning community.	The teacher participates in professional community activities as required, maintaining cordial relationships with colleagues.	The teacher actively participates in the professional community by developing positive and productive professional relationships with colleagues.	*All of level 3 and...* The teacher assumes appropriate leadership roles and promoting positive and professional relationships
	B) Professional development	The teacher resists applying learning gained from professional development activities, and does not share knowledge with colleagues.	The teacher applies learning gained from professional development activities, and makes limited contributions to others or the profession.	The teacher welcomes professional development opportunities and applies the learning gained to practice based on an individual assessment of need. The teacher willingly shares expertise with others.	*All of level 3 and...* The teacher seeks out professional development opportunities and initiates activities that contribute to the profession.
	C) Shared commitment	The teacher demonstrates little commitment to supporting shared agreements that support student learning.	The teacher adheres to shared agreements that support student learning.	The teacher contributes to and actively endorses shared agreements that support student learning.	*All of level 3 and...* The teacher assumes a leadership role in contributing to, endorsing and encouraging others to embrace the shared agreements that support student learning.

Domain 5: *Developing Partnerships with Family and Community*

Standard	Indicators	Does Not Meet Standard Level I	Partially Meets Standard Level II	Meets Standard Level III	Exemplifies Standard Level IV
5.1 Develop two-way communication with families about student learning and achievement	A) Initiation of meaningful communication	The teacher provides minimal information to parents about individual students, and/or the communication is inappropriate to the cultures of the families.	The teacher adheres to the school's required procedures for communicating with families with an awareness of cultural norms	The teacher initiates communication with parents about students' progress on a regular basis, respecting cultural norms.	*All of level 3 and…* The teacher promotes frequent two-way communication with parents to improve student learning with students contributing to the design of the system.
	B) Responsiveness to parent inquiries and communication	The teacher does not respond, or regularly responds insensitively to parent concerns about students.	The teacher responds to parent concerns in a superficial or cursory manner, or responses may reflect occasional insensitivity	The teacher responds to parent concerns in a timely and culturally respectful manner.	*All of level 3 and…* The teacher handles this communication with professional and cultural sensitivity.
	C) Inclusion of the family as a partner in learning decisions	The teacher makes no attempt to engage families in the instructional program, or such efforts are inappropriate.	The teacher makes modest and partially successful attempts to engage families in the instructional program.	The teacher's efforts to engage families in the instructional program are frequent and successful.	*All of level 3 and…* Students contribute ideas for projects that will be enhanced by family participation.
5.2 Equip families with a variety of strategies to support their child's success and college readiness	A) Provision of parent education efforts to support students	The teacher does not provide parents with strategies to support their child's success and college-readiness.	The teacher provides parents with limited strategies to support their child's success and college-readiness.	The teacher provides parents with several strategies to support their child's success and college-readiness including resources outside of the school.	The teacher works collaboratively with parents to identify appropriate strategies to support their child's success and college- readiness including resources outside of the school. Students initiate the use of strategies with their parents.
5.3 Help students leverage resources in their community that support their success in college and beyond	A) Goal setting and advocacy	There is little / no evidence that students work with the teacher to establish learning goals, or that the teacher advocates for students to establish high learning goals.	There is evidence that the teacher advocates for groups of students to establish high learning goals, and that he/she works with students as a group to set goals.	The teacher encourages and advocates for students to attain high learning goals, works to help set and monitor goals, and integrates curriculum experiences that connect to student goals.	*All of level 3 and…* The teacher establishes processes through which students establish and monitor high personal learning goals, and self-advocate for their attainment of the goals.
	B) Knowledge of community resources	The teacher is unaware of resources for students available through the school, CMO or community that students may access to learn about success in college and beyond.	The teacher demonstrates knowledge of resources for students available through the school or CMO, but has limited knowledge of resources available more broadly, or does not work to utilize the available resources to support student understanding of success in college and beyond.	The teacher displays awareness of resources for students available through the school or CMO, and familiarity with resources external to the school and on the Internet; available resources are utilized to increase relevance and student understanding of success in college and beyond.	*All of level 3 and…* Students identify and incorporate resources relevant to them, and that increases their understanding of success in college and beyond.

Domain 5: *Developing Partnerships with Family and Community*

Standard	Indicators	Does Not Meet Standard Level I	Partially Meets Standard Level II	Meets Standard Level III	Exemplifies Standard Level IV
	C) Support for students in accessing these resources	The teacher is unaware of resources and therefore unable to support students accessing resources.	The teacher refers students to other adults in the school to support students in accessing resources.	The teacher supports and advocates for students in accessing resources within and outside of the school by providing information and facilitating personal contacts.	*All of level 3 and...* The teacher promotes the students in taking responsibility for identifying and maintaining contacts with resources.