An Analysis of the Relationship between Teacher Candidate Performance Assessment Results and Teaching Quality at the End of the First Year of Teaching

Joel J. Traver
St. Cloud State University

Follow this and additional works at: http://repository.stcloudstate.edu/edad_etds

Part of the Educational Leadership Commons

Recommended Citation
http://repository.stcloudstate.edu/edad_etds/24

This Dissertation is brought to you for free and open access by the Department of Educational Leadership and Higher Education at theRepository at St. Cloud State. It has been accepted for inclusion in Culminating Projects in Education Administration and Leadership by an authorized administrator of theRepository at St. Cloud State. For more information, please contact kewing@stcloudstate.edu.
An Analysis of the Relationship between Teacher Candidate Performance Assessment Results and Teaching Quality at the End of the First Year of Teaching

by

Joel J. Traver

A Dissertation

Submitted to the Graduate Faculty of

St. Cloud State University

in Partial Fulfillment of the Requirements

for the Degree of

Doctor of Education

in Educational Administration and Leadership

May, 2017

Dissertation Committee:
Dr. John Eller, Chairperson
  Dr. John Hoover
  Dr. Wendy Shannon
  Dr. Roger Worner
Abstract

Nearly a decade after the 1983 landmark report *A Nation at Risk: The Imperative for Education Reform*, the National Commission on Teaching and America’s Future (NCTAF) initiated action to reform education in the United States and established five policy recommendations “aimed at ensuring powerful teaching and learning in all communities as America’s schools and children enter the 21st century” (NCTAF, 1996, p. 4). More specifically, NCTAF sought to reform teacher preparation, professional development, and teacher recruitment with the essential outcome of improved teaching quality and accountability (NCTAF, 1996). Increased teaching quality may have resulted in improved student learning (Cochran-Smith & Villegas, 2015).

The national discourse on accountability of teacher preparation called for both teacher candidate essential knowledge and increased experiences in K-12 schools prior to entering the field as a new teacher (Hollins, 2011). The need for valid and reliable performance assessments, as a part of a system of multiple measures of teaching quality, provide critical information to preparation programs and state-level licensure officers has dominated accountability conversations (Darling-Hammond, 2010). Performance assessments such as the edTPA—developed collaboratively by teachers and teacher educators—represented the complexity of teaching and offered standards that have defined an expert profession (Darling-Hammond & Hyler, 2013).

Darling-Hammond (2010) noted, “[structured teacher performance assessments] have been found to be stronger predictors of teachers’ contributions to student learning gains than traditional teacher tests” (p.7). Possessing the ability to predict teaching quality in the profession by a teacher performance assessment during the pre-service phase of preparation offers exciting possibilities for the field. However, after an extensive review of the literature, no empirical research was discovered on the predictive validity of teaching quality by the edTPA.

Accompanying the edTPA as a measure of teaching quality in Minnesota, several institutions of higher education (IHE’s) in the Midwest have developed and implemented common metrics to evaluate, analyze, and improve teacher preparation programs. Specific to evaluating teaching quality in the field, the Common Metrics Supervisor Survey (CMSS) is administered to obtain a supervisor’s perception of a new teacher’s quality during the first year of teaching. Thus, in order to understand the relationship between the two valid and reliable measures of teaching quality, the study seeks to determine if the edTPA scores are predictive of CMSS results at the end of the first year of teaching.
Acknowledgement

Throughout my experience in St. Cloud State University’s (SCSU) K-12 Educational Administration and Leadership doctoral program, I have received a tremendous amount of encouragement and support from the SCSU faculty and my fellow Cohort VI members. The lifelong friendships and professional connections I have made mean so much to me personally and professionally. I am a better husband, father, and educator as a result.

I would like to extend a special thank you to my dissertation committee comprised of Dr. John Eller, Dr. John Hoover, Dr. Wendy Shannon (WSU), and Dr. Roger Worner. I am forever grateful for your guidance and encouragement throughout this journey. I also want to thank my colleagues at Winona State University who continually offered encouragement and, most importantly, posed challenging questions to improve my study.

I wish to thank my immediate family—my father Guy Traver, mother LaDona Kollasch, and brother Collan Traver and his family—for offering encouragement and exhibiting patience and understanding as I worked to manage family and professional responsibilities.

To my children, Mae Elizabeth and “Baby T”: being your father is the greatest honor I could achieve. I am so excited for our future together. Finally and most importantly, I wish thank my wife, Katie—your unwavering love, support, and sacrifice provided the ability for me to focus on the dissertation journey. Words cannot express how much I love you.
# Table of Contents

List of Tables .................................................................................................................. 7

List of Figures .................................................................................................................. 8

Chapter

I. Introduction ......................................................................................................................... 9
   Statement of the Problem .................................................................................................. 11
   Purpose of the Study ....................................................................................................... 12
   Research Questions ...................................................................................................... 13
   Delimitations .................................................................................................................. 13
   Significance of the Study ............................................................................................... 14
   Definition of Terms ....................................................................................................... 15
   Summary ........................................................................................................................ 17

II. Review of the Literature.................................................................................................. 19
   Purpose of the Review ................................................................................................... 19
   Included in the Review .................................................................................................. 19
   Review of the Related Literature ............................................................................... 19
   Teaching Quality .......................................................................................................... 27
   Teacher Performance Assessments .............................................................................. 29
   Synthesis of the Review of the Research ................................................................... 38
   Summary ........................................................................................................................ 39

III. Method .......................................................................................................................... 40
   Introduction .................................................................................................................... 40
Chapter 1: Purpose of the Study ................................................................. 41
Chapter 2: Research Questions .............................................................. 41
Chapter 3: Research Method ................................................................. 42
Chapter 4: Rationale for Secondary Data Analysis .................................. 42
Chapter 5: Research Design ................................................................. 43
Chapter 6: Rationale for Selection ......................................................... 43
Chapter 7: Population and Sample of Participants .................................. 44
Chapter 8: Instrumentation ................................................................. 45
Chapter 9: Data Collection Procedures ................................................ 52
Chapter 10: Data Analysis ................................................................. 52
Chapter 11: Data Collection Procedures .............................................. 54
Chapter 12: Data Analysis ................................................................. 55
Chapter 13: Limitations ................................................................. 55
Chapter 14: Summary ................................................................. 55

IV. Results .................................................................................. 56

Chapter 15: Introduction ................................................................. 56
Chapter 16: Research Problem/Purpose .............................................. 58
Chapter 17: Research Questions ......................................................... 58
Chapter 18: Method ........................................................................ 59
Chapter 19: Results for Research Questions ......................................... 63
Chapter 20: Synthesis ................................................................. 80
Chapter 21: Summary ................................................................. 81
Chapter V. Discussion

Introduction .................................................................................................................................................................................. 82
Discussion and Conclusions .......................................................................................................................................................... 85
Limitations ................................................................................................................................................................................... 89
Recommendations for the Field ..................................................................................................................................................... 90
Recommendations for Future Research ................................................................................................................................... 91
Summary ..................................................................................................................................................................................... 92

References ................................................................................................................................................................................. 94

Appendices

A. St. Cloud State IRB Approval Form ....................................................................................................................................... 101
B. edTPA Alignment to InTASC Core Teaching Standards ...................................................................................................... 102
C. Common Metrics Supervisor Survey Alignment to InTASC Standards ................................................................................... 103
D. edTPA Reliability .................................................................................................................................................................... 105
# List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Cut Scores and Overall Passing Rates</td>
<td>36</td>
</tr>
<tr>
<td>2. edTPA and CMSS Alignment to InTASC Core Teaching Standards</td>
<td>57</td>
</tr>
<tr>
<td>3. edTPA Descriptive Statistics Based on Pearson Education Reports</td>
<td>65</td>
</tr>
<tr>
<td>4. edTPA Descriptive Statistics Based on InTASC Standard Categories</td>
<td>66</td>
</tr>
<tr>
<td>5. CMSS Descriptive Statistics Based on InTASC Standard Categories</td>
<td>67</td>
</tr>
<tr>
<td>6. Linear Regression: Use Overall edTPA Score to Predict CMSS Overall Average Score</td>
<td>69</td>
</tr>
<tr>
<td>7. Linear Regression (#1): Use edTPA InTASC Category 1: The Learning and Learning Scores to Predict CMSS InTASC Category 1: The Learner and Learning Average Scores</td>
<td>72</td>
</tr>
<tr>
<td>8. Linear Regression (#2): Use edTPA InTASC Category 2: Content Scores to Predict CMSS InTASC Category 2: Content Average Scores</td>
<td>74</td>
</tr>
<tr>
<td>9. Linear Regression (#3): Use edTPA InTASC Category 3: Instructional Practice Scores to Predict CMSS InTASC Category 3: Instructional Practice Average Score</td>
<td>76</td>
</tr>
<tr>
<td>10. Linear Regression (#4): Use edTPA InTASC Category 4: Professional Responsibility Scores to Predict CMSS InTASC Category 4: Professional Responsibility Average Scores</td>
<td>78</td>
</tr>
</tbody>
</table>
List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. InTASC Model Core Teaching Standards Categories</td>
<td>56</td>
</tr>
<tr>
<td>2. edTPA Scores and CMSS Overall Average Score</td>
<td>70</td>
</tr>
<tr>
<td>3. edTPA Category 1: The Learning and Learning Scores and CMSS Category 1: The Learning and Learning Scores</td>
<td>73</td>
</tr>
<tr>
<td>4. edTPA Category 2: Content Scores and CMSS Category 2: Content Scores</td>
<td>75</td>
</tr>
<tr>
<td>5. edTPA Category 3: Instructional Practice Scores and CMSS Category 3: Instructional Practice Scores</td>
<td>77</td>
</tr>
<tr>
<td>6. edTPA Category 4: Professional Responsibility Scores and CMSS Category 4: Professional Responsibility Scores</td>
<td>79</td>
</tr>
</tbody>
</table>
Chapter I: Introduction

Ensuring American public school students learn from high-quality teachers is of critical importance for the continuance of American democracy (White, Van Scotter, Haroonian, & Davis, 2007). The most impactful school-based factor affecting student learning is the quality of teaching in the classroom (Goldhaber, 2007; McCaffrey, Lockwood, Koretz, & Hamilton, 2003; Rivkin, Hanushek, & Kain, 2000; Rockoff, 2004; Rowan, Correnti & Miller, 2002; Wright, Horn, & Sanders, 1997). Several researchers have identified teacher effectiveness, teacher quality, or teaching quality as an effect on student growth of at least 20 percentile points. For example, Goldhaber’s (2007) large-scale study revealed an effect size of 1.4 attributed to teacher quality. Rivkin et al. (2000) discovered an effect size of 0.6-0.7, while Rockoff (2004) revealed an effect size of 0.6.

As a result of these studies, the national political debate on the quality of teaching in American public schools that began in the 1960s continues to the present (Cochran-Smith & Fries, 2001; Wiseman, 2012). More specifically, the use of public funds for teacher preparation and perceived lack of teacher quality has led to renewed efforts of accountability for states and colleges of teacher education across the United States (Lewis & Young, 2013). “States must assure families that teachers meet a standard of competence because parents are compelled by law to entrust their children to educators in schools” (Darling-Hammond & Hyler, 2013, p. 14). The Minnesota Board of Teaching (BOT) administers the accountability system of teacher preparation programs in the state of Minnesota by establishing standards on multiple measures (Minnesota Association of Colleges of Teacher Education [MACTE], 2015).

The Minnesota BOT was established in 1967 by the Legislature to establish and maintain licensure standards and requirements, approve institutions and licensure programs to prepare teachers, and establish and enforce the educator’s Code of Ethics (MACTE, 2015). Part of a system of multiple measures of teacher preparation program effectiveness, the Minnesota Legislature passed legislation requiring the BOT to establish rules requiring all Minnesota teacher candidates to complete a performance assessment in 2011 (edTPA Minnesota, n.d.-b). To enact the legislation, the BOT established rigorous, multiple measures to assess teacher preparation standards set by the Minnesota Legislature and the BOT to grant approval of teacher preparation programs (MACTE, 2015). “The purpose of the program and unit approval process is to guarantee that ALL approved programs demonstrate the required, high results expected in Minnesota” (MACTE, 2015, p. 3).

The Minnesota Association of Colleges of Teacher Education (MACTE) and its member institutions adopted the Education Teacher Performance Assessment (edTPA) to meet state accreditation standards as determined by the Minnesota BOT (Minnesota Association of Colleges of Teacher Education [MACTE], 2013). The edTPA is a nationally implemented valid and reliable assessment of teaching quality completed during the teacher candidate’s student teaching semester (Darling-Hammond, 2010). During the 2012-2013 academic year, all 31 Minnesota teacher preparation institutions began requiring teacher candidates to complete the edTPA (edTPA Minnesota, n.d.-b). In order to meet Minnesota BOT accreditation standards, teacher preparation programs must analyze candidate results and utilize the data for program improvement (edTPA Minnesota, n.d.-b).

In September of 2014, the Minnesota BOT established the following “cut scores” on the edTPA to evaluate teacher preparation program effectiveness:
70% of program completers must earn a score of 13 on Task 1: Planning; 70% of program completers must earn a score of 13 on Task 2: Instruction; and 70% of program completers must earn a score of 12 on Task 3: Assessment. Percentages of program completers to earn specific aggregate scores have not been determined. (Minnesota Board of Teaching [BOT], 2014)

Minnesota IHE’s have only fully implemented the edTPA since the 2013-2014 academic year, and the impacts of edTPA results must be studied locally. It is known that Minnesota teacher preparation program representatives must analyze and utilize edTPA results for program improvement and accreditation (MACTE, 2013). On the other hand, it is unknown if educational leaders in K-12 public schools analyze or utilize teacher candidate edTPA scores, understand how to interpret edTPA scores, or if the edTPA is predictive of teaching quality. Understanding to what extent a teacher candidate’s edTPA results during student teaching are related to teaching quality during a new teacher’s first year is of great importance to educational leaders in both K-12 school districts and higher education.

**Statement of the Problem**

After an extensive review of the literature, no empirical research was discovered on the predictive validity of teaching quality by the edTPA. Newton’s (2010) study claimed to validate the edTPA’s predictive ability for teacher effectiveness based upon Value-Added Modeling (VAM). VAM is a statistical approach for estimating effects of individual teachers or schools on student achievement while accounting for differences in student background, such as low socio-economic status, race, gender, etc. (American Statistical Association, 2014). Educational researchers including Linda Darling-Hammond (2015) critiqued the use of VAM to predict and determine teacher effectiveness. As a result of the lack of evidence supporting the predictive ability of teacher effectiveness of the edTPA, Stanford Center for Assessment, Learning, and Equity (SCALE), publisher of the edTPA, released a statement claiming that predictive validity
of teacher effectiveness cannot be fully achieved by the edTPA because the student teaching experience is too short of a timeframe to collect data (Darling-Hammond, 2015).

In a working paper by the National Center for the Analysis of Longitudinal Data in Education Research (CALDER, 2016), authors Goldhaber, Cowan, and Theobald (2016) reported that the relationship between edTPA scores and teaching effectiveness is mixed. According to Goldhaber et al. (2016): “When we consider that the edTPA is a binary screen of teaching effectiveness (i.e., pass/fail), we find that passing the edTPA is significantly predictive of teacher effectiveness in reading but not in mathematics” (p. iii).

Newton’s (2010) claims of predictive validity based on VAM, as well as the Goldhaber et al. (2016) research, based their findings on edTPA in relation to teacher effectiveness—not teaching quality. As a result, insufficient research exists on the edTPA’s predictive ability of teaching quality. The study seeks to examine the relationship between selected teacher candidate edTPA scores and perceived teaching quality after the first year of teaching.

**Purpose of the Study**

The purpose of the study is to examine the relationship between edTPA scores and perceived teaching quality after the first year of teaching. According to Darling-Hammond (2010), performance assessments in the pre-service stage of preparing educators (such as the edTPA) could help answer questions regarding predictive metrics of teaching quality. Darling-Hammond (2010) asserted that, “[structured teacher performance assessments] have been found to be stronger predictors of teachers’ contributions to student learning gains than traditional teacher tests” (p. 7). Accompanying the edTPA as a measure of teaching quality in Minnesota, several institutions of higher education (IHE’s) in the Midwest have developed and implemented common metrics to evaluate, analyze, and improve teacher preparation programs. Specific to
evaluating teaching quality in the field, the Common Metrics Supervisor Survey (CMSS) asks a new teacher’s supervisor their perception of that teacher’s quality during the first year of teaching. At present, the instruments—edTPA and CMSS—are used separately by IHE’s. The study intends to analyze data collected by the two instruments to ascertain whether or not a relationship exists between them.

**Research Questions**

Based upon the assertion that teaching quality is the most impactful school-based factor on student learning (McCaffrey et al., 2003; Rivkin et al., 2000; Rowan et al., 2002; Wright et al., 1997), and the presence of two existing data sets on Minnesota teacher candidate teaching quality—edTPA and CMSS—the following research questions will guide the study:

1. To what extent do overall edTPA scores predict teaching quality at the end of the first year as determined by supervisor responses on the Common Metrics Supervisor Survey?

2. To what extent do edTPA scores predict CMSS responses related to “InTASC Category 1: The Learner and Learning?”

3. To what extent do edTPA scores predict CMSS responses related to “InTASC Category 2: Content?”

4. To what extent do edTPA scores predict CMSS responses related to “InTASC Category 3: Instructional Practice?”

5. To what extent do edTPA scores predict CMSS responses related to “InTASC Category 4: Professional Responsibility?”

To answer the research questions, a criterion-referenced predictive validity study will be conducted.

**Delimitations**

According to Roberts (2010), delimitations are the intentional narrowing of the study to identify the limits, or boundaries, of the study in order to describe aspects of the study that are
intentionally included or excluded. Delimitations are aspects of a study that can be controlled, such as methods and sample (Simon, 2011). The researcher has established the following delimitations—Timeframe, Sample, and Criterion,

**Timeframe.** The edTPA and CMSS are administered between six and 12 months apart (depending on the teacher candidate’s graduation). In other words, respondent’s timely response to requests for CMSS completion are a primary delimiting factor. Existing edTPA data will be analyzed that were collected during the 2013-2016 academic years (AY). Existing Common Metrics Supervisor Survey (CMSS) data will be analyzed that were collected from May 2014-July 2014, May 2015-July 2015, and May 2016-July 2016 by a small public university in Minnesota. So, the edTPA data set could have contained fall and spring administrations for the academic years of 2013, 2014, 2015, as well as fall 2016. The CMSS is administered several months after edTPA, thus excluding 2016 data, and timeliness of responses may have limited the number of cases included in the study.

**Sample.** Data will be examined on 120 teacher candidates prepared at a small public university in Minnesota who were subsequently hired into teaching positions.

**Criterion.** To be included in the study, teacher candidates must have a scored edTPA, must have been hired into a teaching position and had their supervisors submit an evaluation, or perception, of the new teacher’s teaching quality to the university through the CMSS. The researcher will match the CMSS responses to candidate edTPA scores for statistical analysis.

Further, little research has been conducted on the CMSS as a criterion variable.

**Significance of the Study**

The edTPA has not been studied in the state of Minnesota in terms of the predictive ability of teaching quality when compared to a new teacher’s supervisor’s perception of teaching
quality. The only study concerning Minnesota teacher candidate involvement with the edTPA was conducted during the pilot phase in 2012 regarding candidate understanding of scientific inquiry as required by the science content-specific edTPA. The study aims to examine the relationship between edTPA scores and perceived teaching quality at the end of the first year of teaching. The findings of the study will be useful to Minnesota educational leaders in both K-12 school districts and higher education institutions because if it is determined there is a positive relationship between edTPA scores and teaching quality, the predictive ability of the edTPA will be confirmed. If that outcome is confirmed, the edTPA will continue to be valued for the high-stakes evaluation of teacher preparation programs and for use by K-12 educational leaders in assessing the quality of teacher candidates during the hiring process.

Definition of Terms

Locke, Spirduso, and Silverman (2007) describe the definition of terms as an author’s use of defined terms to prospectively provide readers outside of the field of study a common and understandable language. According to Roberts (2010), the definition of terms provides operational definitions for terms that do not have a commonly known meaning within the research study. The study uses the following terms:

Common Metrics Supervisor Survey: A quantitative survey asking a supervisor to assess the quality of a selected first-year teacher’s instructional practices, abilities to work with diverse learners, abilities to establish positive classroom environments, and levels of professionalism (Hezel Associates, 2013).

Criterion-referenced, concurrent validity: “A test relating the test scores of a group of subjects to a criterion measure standardized test administered at the same time or within a short interval of time” (Borg & Gall, 1989, p. 254).
Criterion-referenced, predictive validity: “A test relating the test scores of a group of subjects to a criterion measure standardized test administered after a period of several months or more” (Borg & Gall, 1989, p. 254).

Criterion variable: The presumed effect in a non-experimental study (Indiana University, Bloomington, n.d.-a).

edTPA: A valid and reliable performance assessment implemented nationally as a measure of teaching quality of teacher candidates prior to licensure and in-service teaching (Stanford Center for Assessment, Learning, and Equity (SCALE, 2013b).

Effect size: “Identifies the strength of the conclusions about group differences or the relationships among variables in quantitative studies” (Creswell, 2009, p. 229).

InTASC Core Teaching Standards: “Common core principles and foundations of teaching practice that outline what teachers should know and be able to do to ensure every PK-12 student reaches the goal of being ready to enter college or the workforce” (Council of Chief State School Officers [CCSSO], 2013, p. 3).


Performance assessment: “Assessments that require students to craft their own responses to problems through constructing an answer, producing a product, or performing an activity rather than merely selecting from multiple-choice answers” (Parsi & Darling Hammond, 2015, p. 1).
Program completer: A teacher candidate who has successfully met the standards and outcomes of a teacher preparation program, including successful completion of student teaching (Hezel Associates, 2013).

Reliability: “Whether scores to items on an instrument are internally consistent, stable over time, and whether there was consistency in test administration and scoring” (Creswell, 2009, p. 233).

Teacher candidate: A student admitted to a teacher preparation program learning to become an educator and has yet to successfully complete standards and outcomes of their teacher preparation program, including student teaching prior to licensure (Larson, 2010).

Teacher quality: “The personal traits, skills, and understandings an individual brings to teaching, including dispositions to behave in certain ways” (Darling-Hammond, 2007, p. 2).

Teacher effectiveness: The impact of an individual educator’s teaching on student achievement results (Newton, 2010).

Teaching quality: “Strong instruction that enables a wide range of students to learn; instruction meets the demands of the discipline, the goals of instruction, and the needs of students in a particular context” (Darling-Hammond, 2007, p. 4).

Validity: “Whether one can draw meaningful and useful inferences from scores on particular instruments” (Creswell, 2009, p. 235).

Value-added research model: Measures aimed to determine how much of a student’s academic progress from one year to the next is attributable to his or her teacher, as opposed to factors outside of the teacher’s control (Center for Education Policy Research, 2011).

Summary

Chapter I of the study contains an introduction to the study, statement of the problem,
purpose of the study, research questions, significance of the study, delimitations of the study, and definition of terms. Chapter II presents a review of the related literature in relation to accountability of teacher preparation nationally and in Minnesota; the rationale, history, and implementation of the edTPA; and an overview of performance assessments—specifically the edTPA; and teacher retention. Chapter III provides the methodology utilized during the data collection and analysis of the study, including an overview of methods, research design, setting of the study, participant selection, data identification and analysis. Chapter IV details the findings of the study. Finally, Chapter V provides a summary, establishes conclusions, and offers recommendations and suggestions for further research.
Chapter II: Review of the Literature

Purpose of the Review

The purpose of the review of the related literature is to analyze and synthesize current and past research on three emerging themes supported by the study: theoretical historical perspectives of teacher preparation; teaching quality; and teacher performance assessments and the implementation of the edTPA in Minnesota.

Included in the Review

The study is based on a theoretical perspective of Habermas’ (1972, 1974) three knowledge constitutive interests and Schön’s (1983) reflective practice. Application of theory into practice is framed by Shulman’s (1998) conception of teaching as a profession. Included in the review of related literature were articles on theoretical and conceptual understandings of teaching, as well as historical reforms in teacher preparation. Both national and local efforts of teacher preparation reform were discussed. Performance assessments as a means of authentic evaluation of teacher candidates prior to their first year of teaching were also discussed from a national and local level. More specifically, the researcher focuses on a specific performance assessment—the edTPA—as a measure of teaching quality and evaluation of teacher preparation programs in Minnesota.

Review of the Related Literature

Theoretical and conceptual framework. In order to provide boundaries for—or rationalize the study—the researcher sought to ground the study in a theoretical framework (Roberts, 2010). Rennert-Ariev (2005) applied a theoretical framework for grounding authentic assessments of teachers (i.e. the edTPA) in a critical analysis of the emerging trend of authentic assessments of teaching quality. Rennert-Ariev (2005) provided educational leaders a theoretical
understanding regarding the characteristics of authentic assessment in order to justify the implementation and outcomes of these assessments. Rennert-Ariev (2005) applied Habermas’ (1972, 1974) three knowledge constitutive interests—the technical, the practical, and the emancipatory—in the rationalization of authentic assessment of teachers. Habermas’ knowledge constitutive interests was discussed in detail in the review of related literature. Rennert-Ariev (2005) argued, “. . . [if] authentic assessment practices [are] to be implemented in ways that are true to their underlying philosophy, and to avoid their misuse as more bureaucratic testing, [the knowledge constitutive interests] should prove useful for educators and policymakers engaged in the reform of testing practices for teachers” (Rennert-Ariev, 2005, p. 2-3).

Chung (2008) later provided three theoretical frameworks to justify and explain performance assessments in teacher education. Chung discussed Schön’s (1983) theory of “reflection in action” as a guiding principle in explaining how teaching is a process of carrying out an action while problem solving—thus, resulting in professional learning. Chung applied Schön’s theory to the PACT (Performance Assessment for California Teachers)—which later became the edTPA—by explaining how the performance assessment requires a teacher to “examine and reflect on a complete cycle of teaching from planning . . . to evaluating student learning and devising changes in future practice . . . [which] may evoke ‘reflection in action’ . . . [that] underlie the profession” (Chung, 2008, p. 9). Further, in relation to the fact that “contexts in which teachers learn to teach may mediate the extent to which any intervention aimed at improving teachers’ instructional practice can have an impact,” Chung (2008, p. 10) applied “situated knowledge theory” posited by Bruner (1996); Greeno, Collins, and Resnick (1996) as well as “social constructivist theory” postulated by Gage & Berliner (1998).

The idea of a ‘profession’ describes a special set of circumstances for deep understanding, complex practice, ethical conduct, higher-order learning, circumstances that define the complexity of the enterprise and explain the difficulties of prescribing both policies and curriculum in this area. (p. 515).

Furthermore, Shulman (1998) posited six characteristics to describe professions:

- the obligation of service to others, as in a ‘calling;’
- an understanding of a scholarly or theoretical type;
- a domain of skilled performance or practice;
- the exercise of judgment under conditions of unavoidable uncertainty;
- the need for learning from experience as theory and practice interact;
- a professional community to monitor quality and aggregate knowledge. (p. 516)

The theoretical and conceptual underpinnings of teaching described above provide the basis for the analysis of teaching quality and performance assessments within the study.

**Historical synthesis of teacher preparation reform.** Pressure on educational leaders to reform teaching quality and improve student learning in American schools has been important in the field of education since the 1983 National Commission on Excellence in Education’s (NCEE) report *A Nation at Risk* (Lewis & Young, 2013). Over the past 30 years, responsibility for educational outcomes has from national and state governments to colleges and universities, local communities, school boards, administrators, teachers, and students (Lewis & Young, 2013). As low student achievement persisted through the various reforms, the focus on accountability shifted from students to teachers (Lewis & Young, 2013).
As one of the largest professions in the United States, educational reformers and policy makers focus on improving the quality and performance of teachers in an effort to have the greatest impact on students and society in general (Cochran-Smith & Villegas, 2015). To illustrate, nearly a decade after the publishing of *A Nation at Risk*, the National Commission on Teaching and America’s Future (NCTAF) initiated action to reform education in the United States by establishing five policy recommendations “aimed at ensuring powerful teaching and learning in all communities as America’s schools and children enter the 21st century” (NCTAF, 1996, p. 4). With specific emphasis on reforming teacher preparation, NCTAF sought to “reinvent teacher preparation and professional development [and] fix teacher recruitment” (NCTAF, 1996, p. 11). The essential outcome of a reformed system of teacher recruitment and preparation was improved teacher quality, thus resulting in improved student learning (Cochran-Smith & Villegas, 2015).

The NCTAF (1996) report proposed that by the year 2006, each student in the United States would have achieved “his or her birthright: access to competent, caring, qualified teaching in schools organized for success” (p. 10). Calling for greater accountability for teacher preparation programs, NCTAF and other professional organizations led the creation of two sets of professional standards for teacher candidates. The first set of teacher education standards were developed by the Interstate Teacher Assessment and Support Consortium (InTASC) of the Council of Chief State School Officers (2011).

[The standards] outline what teachers should know and be able to do to ensure every K-12 student reaches the goal of being ready to enter college or the workforce in today’s world. These standards outline the common principles and foundations of teaching practice that cut across all subject areas and grade levels and that are necessary to improve student achievement. (p. 3)
The second set of standards for the education profession was developed by an existing professional organization, the National Council for the Accreditation of Teacher Education (NCATE) (Brabeck & Koch, 2013). Empowered by policy and reform advocates, NCATE acted to ensure professional standards were met by teacher preparation programs using program audits and accreditation (Lewis & Young, 2013).

According to Knight et al. (2014), teacher preparation accreditation shifted from two separate sets of professional organizations and standards to a consolidated effort to improve teacher education. To illustrate, NCATE and the Teacher Education Accreditation Council (TEAC) consolidated into the Council for the Accreditation of Educator Preparation (CAEP) as the field’s reputable teacher preparation program accrediting body. As a result of the consolidation into CAEP, as well as the emphasis to include performance assessments in the evaluation of programs, CAEP initiated the Blue Ribbon Panel (BRP) on clinical preparation. The BRP was established to work with the Teacher Performance Assessment Initiative and the American Association of Colleges for Teacher Education (AACTE) member institutions to respond to the political and professional calls for authentic assessments of teacher candidates (Knight et al., 2014). As a result, the work of the BRP sought to increase teaching quality in the field (Knight et al., 2014).

In a synthesized analysis of literature of reform of teacher preparation, Cochran-Smith (2001) deduced three primary “outcomes” dominating the literature. Specifically, the three primary “outcomes” consist of: long-term impact, teacher test scores, and professional performance (Cochran-Smith, 2001, p. 527). Long-term impact includes the key indicators of teacher knowledge, preparedness, teacher attrition, ratings, and student achievement (Cochran-Smith, 2001). Teacher test scores measure basic skills in reading, writing, mathematics, content
knowledge and pedagogy knowledge assess a candidate’s “fit” for the profession (Cochran-Smith, 2001). Professional performance included teacher-produced materials such as portfolios or authentic assessment tasks (such as the edTPA) (Cochran-Smith, 2001). As a result of her analysis, Cochran-Smith (2001) called for the education profession to determine the measures of accountability that will initiate and sustain effective educational reform for improved teaching and learning, and professors, teachers, and educational leaders must develop those measures themselves. “Unless we keep at the center of teacher education rich and complex understandings of teaching and learning . . . not easily reducible to algorithms, [teachers are] neither the saviors nor the culprits in what is wrong with American schools and American society” (Cochran-Smith, 2011, p. 543-544). Darling-Hammond and Hyler (2013) contend that the authentic evaluation of teaching quality using performance assessments answers the call for accountability. Performance assessments such as the edTPA—developed collaboratively by teachers and teacher educators—represents the complexity of teaching and offer standards that can define an expert profession (Darling-Hammond & Hyler, 2013).

**Historical synthesis of teacher preparation reform in Minnesota.** The state requirement for the Minnesota Board of Teaching (BOT) to adopt a performance assessment for teacher candidates in conjunction with the Minnesota Association of College of Teacher Education (MACTE), was approved by the Minnesota legislature and signed by Minnesota’s Governor in 2011 (edTPA Minnesota, n.d.-b). The applicable section was added to a list of duties assigned to the BOT and reads as follows (edTPA Minnesota, n.d.-a):

(d) The Board must provide the leadership and adopt rules for the redesign of teacher education programs to implement a research based, results-oriented curriculum that focuses on the skills teachers need in order to be effective. The board shall implement new systems of teacher preparation program evaluation to assure program effectiveness based on proficiency of graduates in demonstrating attainment of program outcomes.
Teacher preparation programs including alternative teacher preparation programs under section 122A.245, among other programs, must include a content-specific, board-approved, performance-based assessment that measures teacher candidates in three areas: planning for instruction and assessment; engaging students and supporting learning; and assessing student learning.

In 2011, the above statute passed by the Minnesota Legislature required teacher preparation programs in Minnesota to include a BOT-approved performance assessment as part of the preparation of the state’s teachers (MACTE, 2013). However, prior to this formal legislation, the Minnesota BOT and MACTE had begun to discuss how to include performance assessments in the state’s accreditation system for teacher preparation programs (MACTE, 2013). As part of that process, Minnesota joined IHE’s across the nation as the SCALE and leading experts in the field developed and piloted the edTPA (formerly known as TPA) (edTPA Minnesota, n.d.-b).

In the fall of 2010, before the legislature placed the performance assessment requirement in law, the BOT and MACTE conducted the first edTPA Implementation Summit, established the edTPA Steering Committee and then hired an edTPA coordinator using grant funds provided by the Archibald Bush Foundation (edTPA Minnesota, n.d.-b). After the new law was enacted, the BOT formally adopted edTPA in 2011 as the statewide performance assessment tool that would be used to meet the state requirement within the new accreditation process (MACTE, 2013). Thus, the cooperation between BOT and MACTE resulted in the edTPA being utilized for teacher preparation program evaluation only and that BOT and the Minnesota Department of Education (MDE) would not be using the edTPA as an evaluation of teacher candidate’s readiness to teach or to earn a teaching license (MACTE, 2013).

In 2012, the Minnesota legislature asked for recommendations from the BOT “for eliminating lower-priority tests or assessments to offset the additional fees charged to students

**NExT and Common Metrics Group.** Educational leaders, faculty, and staff among 14 Institutions of Higher Education (IHEs) in Minnesota, North Dakota, and South Dakota partnered with the Archibald Bush Foundation (non-profit organization) to form the Network for Excellence in Teaching (NExT). The NExT partnership was organized to improve how university-based teacher education programs prepare new teachers. Two sub-groups emerged out of the larger partnership: Common Metrics working group and the Teacher Effectiveness working group.

Members of the Common Metrics group collaborated to develop and administer a set of four common surveys to measure the IHE’s progress toward improving teacher preparation programs and new teacher effectiveness in the field (Hezel Associates, 2013). Teacher candidates and program completers at each of the IHEs complete three surveys upon entry into the teacher education programs, when they complete and exit the program, and one year after graduation (known as the Transition to Teaching Survey or TTS). Supervisors of NExT program completers working in the teaching field complete the Supervisor Survey (CMSS) after the graduate’s first year of teaching.

The CMSS assesses educational leader perceptions of a new teacher’s quality after the first year of teaching in the profession. The CMSS asks supervisors to assess the quality of
graduates’ instructional practices, abilities to work with diverse learners, abilities to establish positive classroom environments, and levels of professionalism. The survey is administered to direct supervisors of teacher education graduates employed in the schools approximately one year after the teachers completed their preparation programs.

**CMSS validity and reliability.** A third-party contractor, Hezel Associates, conducted a validity study of the CMSS in 2012, as well as a reliability study utilizing factor analysis in 2013. Results of the validity study identified that survey items are valid (Hezel Associates, 2012). Results of the reliability study revealed that the 2013 Supervisor Survey had “relatively good reliability” as a measure of evaluating a new teacher’s quality of instructional practice, understanding and work with diverse learners, quality of their classroom environment, and attending to professional responsibilities (Hezel Associates, 2013). As a result of a factor analysis conducted by Hezel Associates (2013), a working group of NExT representatives revised the CMSS prior to the 2014 administration of the survey, so 2013 results are not comparable for most items (Hezel Associates, 2013).

**Teaching Quality**

The most important school-based factor affecting student achievement is the quality of teaching (McCaffrey et al., 2003; Rivkin et al., 2000; Rowan et al., 2002; Wright et al., 1997). As a result, utilizing performance assessment data to quantify teaching quality is important for educational leaders to understand, particularly prior to new teachers entering the classroom.

For the purposes of the study, it is important to differentiate and clearly define three common summations of effective pedagogy: teacher quality, teacher effectiveness, and teaching quality (Darling-Hammond, 2007). Several authors have written about the important differences among the aforementioned descriptors, with specific attention focused on how each evaluation of
teachers is measured within a specific context of students, subject area, and location (Campbell, Kyriakides, Muijs, & Robinson, 2004; Darling-Hammond, 2007).

In the broadest sense, teacher quality is a set of personal characteristics of the educator (Knight et al., 2015). “Teacher quality is the broadest of the three concepts, and refers to the ‘bundle of personal traits, skills, and understandings’ an individual brings to teaching, including dispositions to behave in certain ways” (Darling-Hammond, 2007, p. 2). Newton (2010) added that teacher quality is composed of a set of qualities that would typically be relevant in a variety of teaching contexts.

Teacher effectiveness is not assessed by teacher knowledge, actions, or skills, but by the results of student learning outcomes based on the teaching (Newton, 2010). “Empirically, the primary challenge in assessing teacher effectiveness is the problem of isolating the contribution of the teacher to the outcomes by factoring out other influences that might affect a student’s outcomes” (Newton, 2010, p. 5).

Lastly, teaching quality refers to the quality of a teacher’s work in a particular context, with particular students, in a particular location (Darling-Hammond, 2007). With these specified characteristics, teaching quality is defined more narrowly than teacher quality that involves contextualized practice to address specific conditions (Newton, 2010). More specifically, Darling-Hammond (2007) asserted, “Teaching quality has to do with strong instruction that enables a wide range of students to learn. Such instruction meets the demands of the discipline, the goals of instruction, and the needs of students in a particular context” (p. 5). Adding to the overview of teaching quality, Darling-Hammond (2007) said that collecting data on practice with actual students—the underlying conception of a performance assessment—best assesses teaching quality:
... It is dependent to some extent on the conditions of teaching such as curriculum materials, necessary supplies and equipment, reasonable class sizes, and the opportunity to plan with other teachers to create both appropriate lessons and a coherent curriculum across grades and subject areas. (p. 5)

Thus, given controlled conditions of the context in which teaching occurs, the curriculum being taught, and a specific set of students, performance assessments provide a method of measuring teaching quality (Newton, 2010). “Quality teaching practices involves a transition to a more complex view of teaching behaviors associated with complex learning outcomes and tied explicitly to content” (Knight et al., 2014, p. 1). As a result of the challenge of isolating the factor(s) of teaching, the study focused on teaching quality.

**Teacher Performance Assessments**

The national discourse on accountability of teacher preparation calls for both teacher candidate essential knowledge and K-12 experiences prior to entering the field as a new teacher (Hollins, 2011). The need for valid and reliable performance assessments, as a part of a system of multiple measures of teaching quality, to provide critical information to preparation programs and state-level licensure officers dominates accountability conversations (Darling-Hammond, 2010).

In a report for the Center for American Progress, Darling-Hammond (2010) described the ways in which teacher performance assessments can be used to advance the education profession including improved educational policies and teacher credentialing to evaluation of teaching quality and the impact on student learning:

Specifically, [the report] describes the ways in which assessments of teacher performance for licensing and certification can both reflect and predict teachers’ success with children so that they can not only inform personnel decisions, but also leverage improvements in preparation, mentoring, and professional development. It outlines progress in the field of teacher assessment development and discusses policies that could create much greater
leverage on the quality of teacher preparation and teaching than has previously existed in the United States. (p. 1)

According to Darling-Hammond (2010), traditional metrics of assessing teaching quality—from initial licensing by the teacher preparation institutions to in-service teaching during the probationary period toward tenure—must be replaced by research-based metrics that legitimately evaluate performance rather than merely completing a checklist of actions or activities. Darling-Hammond (2010) believed that current measures for evaluating teachers, both during student teaching and in-service teaching, are not linked to an educator’s capacity to teach. Additionally, educational policies at the local, state, and federal levels for defining and measuring teacher quality rely on methods that are “poor predictors of later effectiveness in the classroom” (Darling-Hammond, 2010, p. 2). For example, traditional evaluations of teaching quality of student teachers or licensed educators include general classroom observations by principals who offer little useful feedback, or focus on paper-and-pencil tests administered to teacher candidates which assess basic academic skills (reading, writing, and mathematics) and subject area knowledge (Darling-Hammond, 2010).

Referencing Newton (2010), Knight et al. (2014) discussed how implementation of the edTPA “improve[s] upon value added models (VAM) by providing information on teaching quality, not just teacher quality or teacher effectiveness, and provide information sooner than might be accomplished by use of VAM alone” (p. 372).

Thus, prompted by the research identifying gaps between educational milestones and actual teaching performance, federal and state policy makers have written legislation and implemented measures of teaching performance in an effort to reform ineffective practices (Reagan, Schram, McCurdy, Change, & Evans, 2016).
In a national analysis of state-implemented teacher performance assessments, Darling-Hammond (2010) analyzed data reported by both Connecticut’s (Beginning Educator Support and Training--BEST) and California’s (Performance Assessment for California Teachers--PACT) teacher performance assessment programs. The analysis sought to substantiate a call for action to create a nationally standardized authentic assessment of teacher knowledge, skill, and application necessary for effectiveness in the classroom. Darling-Hammond (2010) analyzed quantitative data—the “scores”—that teacher candidates earned within the assessment, but also presented qualitative data to contextualize and support the notion that with a performance assessment (and the feedback received from the assessment), teacher candidates must demonstrated the ability to change and improve their practice:

Built on the advances made by the National Board [for Teacher Certification] and Connecticut [BEST] assessments, the Performance Assessment for California Teachers has emerged from this reform. Launched in 2002 and now used by 33 teacher education programs in the state, PACT has been shown to be reliable, valid, and a strong lever for improving both teacher competence and program quality. Like the BEST assessment, a preliminary validity study of PACT also found that teachers’ scores on the assessment are positively associated with their value-added effectiveness when they later become full-time teachers. (p. 10)

As a result of the analysis, Darling-Hammond (2010) found that performance assessments led to increased reflective teaching practices of teacher candidates. The author concluded that such increases in practice ought to convince policymakers to implement performance assessments to measure teacher quality nation-wide, versus “bubble tests” or amount of classes or credits as determining factors of teacher effectiveness. “One of the few areas of consensus among education policymakers, practitioners, and the general public today is that improving teacher quality is one of the most direct and promising strategies for improving public education outcomes in the United States” (Darling-Hammond, 2010, p. 1). Assessing
teacher quality, however, has remained unsuccessful in accurately evaluating a teacher’s capacity to teach (Darling-Hammond, 2010). For example, federal, state, and local policies and practices for evaluating teacher quality rely heavily on classroom observations by principals (Darling-Hammond, 2010). Although principals evaluate most teachers, the principal’s approach in evaluating different teachers is not differentiated, nor does the principal typically provide useful feedback for improving as a result of the evaluation (Darling-Hammond, 2010). Discussing the benefits of performance assessments, Darling-Hammond (2010) stated:

One of the more powerful things about a performance assessment of this kind is that it requires teacher candidates to pull together all the many things they are supposed to be learning in courses and clinical experiences—how to diagnose student learning, plan in response to standards, manage and revise instruction, and evaluate the outcomes for student understanding—into a single coherent teaching event. It may be the first and only time in a program that candidates and their instructors can see whether they indeed understand and can apply what they are supposed to be learning. (pp. 18-19)

Accordingly, due to the strength of performance assessments in providing evidence for predicting teachers’ contributions to student learning, researchers have concluded that states and IHE’s must adopt authentic performance assessments as a component of a system of multiple measures of quality of teacher candidates (Darling-Hammond, 2010; Newton, 2010). Pre-service performance assessments require teacher candidates to analyze and synthesize classroom and research-based information and obtain and analyze feedback to improve instruction (Parsi & Darling-Hammond, 2015).

Performance assessments evaluate directly what teachers do in the classroom, and they often incorporate contextualized evidence of student learning that is linked to evidence of the associated teaching efforts. Such assessments have been found to be stronger predictors of teachers’ contributions to student learning gains than traditional teacher tests. (Darling-Hammond, 2010, p. 7)
Furthermore, when teachers and educational leaders utilize performance assessment results to improve teaching practice, both micro and macro levels of the American educational system are improved to educate all students (Parsi & Darling-Hammond, 2015).

The Performance Assessment for California Teachers (PACT) was developed as an instrument for teaching knowledge applied to a teaching situation as an authentic assessment of teaching quality. Five dimensions are included within the assessment: planning, instruction, assessment, reflection, and academic language. Raters of the PACT are trained and audited thus producing high levels of consistency in scoring (Pecheone & Chung, 2006). Pecheone and Chung (2006) determined that the PACT has a high degree of inter-rater reliability, content validity, and some evidence of construct validity. Pecheone and Chung’s research revealed PACT’s validity and reliability as the preceding performance assessment to the edTPA.

In an early study of the PACT, Newton (2010) linked performance assessment scores with four separate Value-Added Model (VAM) estimates for 14 first- and second-year teachers with 259 students in third through sixth grade at an urban school district in California. Newton used the VAM statistical model to analyze student learning gains of teachers who completed the PACT and compared those results with outcomes for similar students in the sample. The study used the PACT administrative database of assessment submissions. Due to issues with following candidates through their first teaching positions, this exploratory study was conducted largely from candidates from a district intern program as well as one candidate from a traditional teacher education program in the same district.

Newton’s (2010) study presented preliminary findings on the relationship between beginning teacher’s scores on the Performance Assessment for California Teachers (PACT) [which evolved into the edTPA national assessment]. In this limited sample of elementary
English language arts, Newton found a relationship of PACT as a positive predictor of teacher effectiveness. According to Newton, students taught by a teacher scoring at the top of the PACT rating scale (44 points out of 75), scored 20 percentile points higher on standardized English language arts assessments than students taught by a teacher receiving the lowest score on the PACT (24 points out of 75) (Newton, 2010).

Performance assessments can provide predictive information about teacher effectiveness when used for veteran teachers taking the National Board for Professional Teachers exam. Such predictions can be statistically significant even when other teacher quality indicators are included in the model. Such a pattern may generalize to preservice teachers such that performance assessments predict future effectiveness. (Newton, 2010, pp. 8-9)

As PACT results indicated increased teaching quality and teacher effectiveness in California, as well as PACT’s alignment with the National Board for Professional Teaching assessment, the SCALE began to increase the standardization and broad implementation of a teacher candidate performance assessment (Darling-Hammond, 2010; Newton, 2010). SCALE revised the PACT to align with the InTASC standards, resulting in the formation of the Education Teacher Performance Assessment (edTPA) (Balmer & Ofstedal, 2012). As part of the “scale up,” 25 states across the United States began to develop and implement performance assessment policy—including Minnesota and Washington as early implementation states of the national edTPA field tests (Balmer & Ofstedal, 2012).

The edTPA, modeled after the valid and reliable PACT (Pecheone & Chung, 2006), began to be widely implemented in states across the nation beginning in 2010 (Sato, 2014). “The edTPA draws from experience gained over a 25-year history developing performance-based assessments of teaching, including the National Board for Professional Teaching Standards and the Performance Assessment for California Teachers (PACT)” (Balmer & Ofstedal, 2012).
SCALE (2013b) conducted an analysis of the early implementation of the edTPA with a national field test. Researchers analyzed over 4,000 teacher candidate edTPA submissions from the 2013 national field test (SCALE, 2013b). The District of Columbia and 29 states, at the time of the study, participated in the edTPA field test. The 2013 edTPA field test evaluated 4,055 submissions across 23 content fields, including secondary social studies, elementary mathematics, and K-12 visual arts for example (SCALE, 2013b). According to SCALE (2013b):

Each submission was scored on five separate rubrics within each of the three major edTPA tasks: planning, instruction, and assessment. With five levels of possible performance for each rubric (and 15 rubrics total for most assessments), the total score could range from 15 to 75. (p. 17)

Trained evaluators score edTPA portfolios within an electronic portfolio management system. The electronic portfolio management systems allow evaluators to view and score candidate materials related to each task of the edTPA performance assessment (SCALE, 2013b). The electronic portfolio management system provides evaluators, SCALE, and Pearson Education to efficiently collect scores assigned to each submission for reporting results for teacher candidates and data analyses for teacher preparation institutions (SCALE, 2013b).

Adding to the validity and reliability of the edTPA discussed in Chapter 3, SCALE (2013b) discussed:

The edTPA [was] developed within a technical framework of psychometric practice and principles guided by The Standards for Educational and Psychological Testing. The edTPA development process yielded important evidence to provide the foundation of the assessments’ validity, reliability, and usability for the purposes of teacher licensure, accreditation of teacher preparation programs, and candidate completion of preparation programs. (p. 17)

SCALE (2013b) conducted several data analyses within the edTPA field test. Researchers conducted validity, content validity, job analyses (i.e. skills and behaviors), bias and sensitivity review, subgroup score analysis, construct validity, reliability, and inter-rater
agreement rates. The report provided stakeholders a summary of the design and development process of the edTPA, as well as validation results for 4,055 teacher candidate submissions.

SCALE (2013b) concluded that the edTPA “provides a measure of teacher candidates’ readiness to teach that can inform program completion, licensure, and accreditation decisions, while supporting candidate learning and preparation program improvement” (p. 1).

As a result of the national field test described above, standard-setting was enacted to identify levels of passing performance, or numerical scores identifying a teacher candidate’s readiness to teach. Passing cut scores between 37-42 points (within one standard error of measurement of the maximum recommended cut score) were analyzed (SCALE, 2013b). The following table identifies cut scores and overall passing rates of the 4,000 submissions:

Table 1

<table>
<thead>
<tr>
<th>Cut Score</th>
<th>Overall Passing Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>37</td>
<td>78.0%</td>
</tr>
<tr>
<td>38</td>
<td>74.3%</td>
</tr>
<tr>
<td>39</td>
<td>70.5%</td>
</tr>
<tr>
<td>40</td>
<td>66.8%</td>
</tr>
<tr>
<td>41</td>
<td>62.1%</td>
</tr>
<tr>
<td>42</td>
<td>57.9%</td>
</tr>
</tbody>
</table>

Note: Data from edTPA Field Study (SCALE, 2013b).

As a result, SCALE (2013b) recommended states establish a cut score ranging between 39-42 points (a half standard error of measurement) to be utilized in the standard-setting process for each state that implements the edTPA.
In the state of Minnesota, as previously discussed, candidate scores on the edTPA are not utilized for licensing or credentialing purposes. Rather, scores are utilized for teacher preparation program assessment (Balmer & Ofstedal, 2012).

In a study of edTPA implementation for the world languages content area, Hildebrandt and Swanson (2014) examined edTPA scores of 21 world language teacher candidates from two teacher preparation programs in Illinois and Georgia and compared those results to the candidate’s scores for the states of Washington and New York.

One teacher preparation program in Illinois and one program in Georgia received candidates’ numerical scores on the 13 rubrics of the Foreign Language edTPA and data were entered into SPSS 19.0 and analyzed. Following the researchers’ analysis of their candidates’ scores, the researchers mined existing data of participants’ submitted edTPA to Pearson Education in a comparison study. Findings concluded that participants scored the highest on Task 1: Planning, and lowest on Task 3: Assessment. Candidate scores in the study ranged from 31-52 and were compared to cut scores in Washington and New York—states that had fully implemented edTPA. Given the available data, all participants would have earned licensure in Washington (cut score of 30) and 90% would have met or exceeded the 35 cut score in New York for licensure (Hildebrandt & Swanson, 2014).

Limited research exists on the implementation and results of edTPA in the state of Minnesota. In an unpublished dissertation for the University of Minnesota, Billington (2012) examined 29 Science teaching candidates at a large midwestern university. The subjects were teacher candidates in a post-baccalaureate teacher licensure program; as such, all 29 subjects had bachelor degrees in the Science content area. The demographic breakdown of subjects equated
Billington (2012) conducted the study during the pilot phase of the edTPA in the state of Minnesota. For the study, Billington established two phases when analyzing Science teacher candidate’s understanding of inquiry-based instruction as assessed by the edTPA:

Phase 1: Analysis of Written Commentary in Task 2 for 26 candidates; Early, Mid, Reflective Journals of 26 candidates; 20 minute video clips in Task 2 for 26 candidates; Phase 2: Semi-structured interviews of 6 candidates; Reflective journals of 6 candidates. (p. 31)

The Billington (2012) study revealed that teacher candidates’ scores on the edTPA were directly related to their ability to hear, understand, and evaluate the content of student conversations. Most candidates had appropriate views of inquiry-based instruction, and revealed strong reflective practice in both their written commentary and their reflective journal postings. However, the candidates struggled to implement inquiry in meaningful ways, and to capture good evidence of inquiry-based instruction and student engagement for Task 2 of the edTPA.

Synthesis of the Review of the Research

Preserving and extending democracy through the public education of American children has been a national priority for two centuries (White et al., 2007). Research illustrated that the higher the quality of teaching in a classroom, the higher the level of student learning (McCaffrey et al., 2003). Accountability of the preparation of American teachers has led to establishment of national standards for teaching (NCTAF, 1996). Evaluation of teacher candidate quality has been implemented for both IHE’s and individual candidates using authentic teacher performance assessments such as the edTPA (Darling-Hammond, 2010). National and state edTPA data are
being collected and analyzed for local and national improvement of teaching quality (MACTE, 2015).

Summary

Described above, a focus on improving teaching quality remains a priority nationally and in Minnesota. Minnesota lawmakers have legislated reforms for teacher preparation, including the use and implementation of performance assessments such as the edTPA, to collect and analyze teacher candidate performance data for improvement. Several regional IHE’s are collecting teaching quality data from supervisors who have hired teacher candidates prepared at Minnesota IHE’s.
Chapter III: Method

Introduction

In a review of the related literature of teacher preparation effectiveness and accountability, the need for standards-based metrics of teaching quality has manifested (Cochran-Smith & Villegas, 2015). Based upon a conceptual understanding that the act of teaching is one of professional practice (Shulman, 1998), quality teaching will depend on educators who practice effective pedagogy (assessing student learning, relating content, and subscribe to life-long learning) upon entry into the profession (Knight et al., 2015). Thus, coordinated teaching quality metrics are needed to identify, justify, and improve the practice of teaching as a profession. The edTPA (an authentic performance assessment) (SCALE, 2013b) and the Common Metrics Supervisor Survey (survey measuring a new teacher’s supervisor’s perception of teaching quality) are existing data collection tools that can address the need to understand and predict teaching quality.

At present, Minnesota teacher candidate scores on the edTPA are used as an evaluation of teacher preparation programs alone. A small number of Minnesota’s public and private colleges and universities have collaboratively developed and implemented common metrics to gather information on teacher preparation programs for improvement. Common Metrics survey data are utilized in isolation as well. Thus, the relationship between edTPA data and data collected by the Common Metrics Supervisor Survey (CMSS) requires research. As a result, the study aims to examine if there is a relationship between the two instruments to better inform educational leaders in both K-12 and higher education regarding teacher candidate-to-new educator teaching quality.
Purpose of the Study

The purpose of the study is to examine the relationship between edTPA scores and perceived teaching quality after the first year of teaching. According to Darling-Hammond (2010), performance assessments in the pre-service stage of preparing educators (such as the edTPA) could help answer questions regarding predictive metrics of teaching quality. Accompanying the edTPA as a measure of teaching quality in Minnesota, several institutions of higher education (IHE’s) in the Midwest have developed and implemented common metrics to evaluate, analyze, and improve teacher preparation programs. Specific to evaluating teaching quality in the field, the Common Metrics Supervisor Survey (CMSS) asks a new teacher’s supervisor their perception of that teacher’s quality during the first year of teaching. Currently, the instruments—edTPA and Common Metrics Supervisor Survey—are used separately by the IHE’s. The study aims to analyze the data collected by the two instruments to examine if a relationship exists between them.

Research Questions

In order to understand the relationship among edTPA, CMSS, and teaching quality, the following research questions were developed to guide the study:

1. To what extent do overall edTPA scores predict teaching quality at the end of the first year as determined by supervisor responses on the Common Metrics Supervisor Survey?

2. To what extent do edTPA scores predict and CMSS responses related to “InTASC Category 1: The Learner and Learning?”

3. To what extent do edTPA scores predict and CMSS responses related to “InTASC Category 2: Content?”

4. To what extent do edTPA scores predict and CMSS responses related to “InTASC Category 3: Instructional Practice?”
5. To what extent do edTPA scores predict and CMSS responses related to “InTASC Category 4: Professional Responsibility?”

Research Method

I will conduct the study using quantitative methods. Specifically, I will utilize secondary data analysis to answer the research questions. According to Boslaugh (2007), secondary data analysis is an analysis of data that has previously been collected and tabulated by other sources. The secondary data sets that will be analyzed are the edTPA teacher performance assessment and the Common Metrics Supervisor Survey. I will acquire permission to analyze the university’s CMSS and compare those data with existing edTPA results for the university’s teacher preparation program completers. More specifically, the university collects teacher candidate edTPA data fall and spring semesters annually, and CMSS data is collected annually May-July. The CMSS assesses a supervisor’s perception of teaching quality of their newly hired candidates prepared at the university—detailed information of the CMSS is discussed below. I will implement regression statistical analyses to answer the research questions.

Rationale for Secondary Data Analysis

According to Koziol and Arthur (n.d.), secondary data analysis may be an effective means of research where primary data collection is too costly or unattainable, and secondary data is available for analysis in order to answer the research questions. An important limitation of secondary data analysis for consideration is the data were possibly collected for different purposes from which the research intends to answer his or her questions, and thus, may not address the needs of the study (Bhattacherjee, 2012).
**Research Design**

The study will utilize secondary data analysis methodology utilizing two data sources—Education Teacher Performance Assessment (edTPA) and Common Metrics Supervisor Survey (CMSS).

The CMSS and edTPA are both aligned to the national InTASC teaching standards (see Appendices B and C). Both CMSS and the edTPA are assessments collected by the small, public university in Minnesota. Data are utilized for non-research purposes at the university, which puts the survey and assessment in the existing data category. The university stores candidate edTPA samples and edTPA results from past semesters. The university also stores past CMSS data. CMSS data is collected with teacher preparation “program completer” names affixed, so CMSS data can be directly matched to the edTPA for comparative descriptive analysis. “Descriptive analysis refers to statistically describing, aggregating, and presenting the constructs of interest or associations between these constructs” (Bhattacherjee, 2012, p. 119).

Using existing data will enable me to conduct the study without seeking consent from the teacher candidates and supervisors, as identifying information will be removed after matching the data from the CMSS and edTPA. Additionally, validity and reliability of the CMSS (discussed below) will be maintained, as I will use results from the actual CMSS instrument, which has gone through reliability and validity testing. I will not revise portions of the CMSS survey, nor will I use individual items pulled from the larger survey. Additionally, I do not plan to recruit participants, as existing candidate data will be analyzed.

**Rationale for Selection**

Quantitative survey methods—as part of the secondary data analysis methodology—was selected for the study to collect responses from principals that have hired candidates prepared at
the small public university. Survey methodology will allow me to increase the potential response rate in an effort to compare data with candidate scores on the edTPA. Currently, all teacher candidates in Minnesota must complete the valid and reliable (discussed below) edTPA assessment of teaching quality, but candidate scores are not being utilized for making human resource decisions by principals.

Population and Sample of Participants

**Education Teacher Performance Assessment (edTPA).** Teacher candidates (also referred to as student teachers in the related literature) are participants in the study who completed the edTPA during the student teaching semester while completing a teacher preparation program. To be included in the study, teacher candidates must have completed their student teaching semester as a student at the small public university in Minnesota.

The number (n) of respondents on the edTPA varies by semester—fall or spring semester—at the university. For AY 2013-2014, 142 (n=142) respondents completed the edTPA. For AY 2014-2015, 223 (n=223) respondents completed the edTPA. For AY 2015-2016, 190 (n=190) respondents completed the edTPA for a total of 555 (n=555) available cases for the study. Teacher candidates who completed the edTPA were placed in public and private K-12 schools in the United States for student teaching—mostly in the Midwest. K-12 school locations ranged among urban, suburban, and rural settings.

Teacher candidates were selected for inclusion in the study as program completers of the selected university. In discussing criteria for inclusion of participants in the study, not all teacher candidates prepared at the university will qualify for inclusion in the study, as a teacher candidates must have received a numerical score on the edTPA. Participants must also have been
hired for a teaching position for the academic year after participants completed student teaching in order for their supervisor to receive and respond to the Common Metrics Supervisor Survey.

**Common Metrics Supervisor Survey (CMSS).** Respondents to the CMSS are supervisors—most typically K-12 school principals and special education directors—who serve as the primary evaluator of teachers prepared at the selected university. Teacher candidates prepared at the selected university will now be referred to in this discussion as new teachers.

The number (n) of respondents on the CMSS for AY 2013-2014 was 79 (n=79). The number of respondents on the CMSS for AY 2014-2015 was 91 (n=91). The number of respondents on the CMSS for AY 2015-2016 was 87 (n=87). Respondents are located in K-12 schools located in the United States and internationally.

K-12 school supervisors respond to the survey as the primary evaluators of educators’ teaching quality. K-12 school supervisors hire, support, retain, or non-renew new teachers’ contracts within the buildings in which they supervise.

Respondents on the CMSS were included in the study if the new teacher identified them as their supervisor on a different Common Metrics survey—the Transition to Teaching Survey (TTS). The new teacher must have provided directory information for the supervisor in order for them to receive a link to the CMSS.

Convenience sampling was used in the study, as the new teachers self-reported their supervisor. A section of the TTS requests information of the new teacher’s supervisor in order for the supervisor to receive the CMSS.

**Instrumentation**

**edTPA.** In order to answer the research questions utilizing existing data sources, the instrumentation used will be discussed below.
“SCALE is the lead developer for edTPA, and Stanford University is the sole owner of edTPA. The university has an agreement with Evaluation Systems, a unit of Pearson Education, licensing it to administer and distribute edTPA” (SCALE, 2013a, Preface and Acknowledgements). The edTPA is an authentic performance assessment to evaluate a teacher candidate’s preparation in one of 27 initial licensure areas. The edTPA requires teacher candidates to systematically examine the cycle of teaching aimed at evaluating specific learning goals, using evidence to demonstrate their planning, instruction, and student assessment (SCALE, 2013b). Administrators of the edTPA collect subject-specific evidence of from a teaching segment of three-to-five lesson plans contained within a unit of instruction in a single classroom context. In addition to lesson plans and supporting materials, teacher candidates submit a video and reflective journals (called commentaries) to justify the pedagogical approaches they chose in the learning segment. Trained evaluators assess the teacher candidate’s work within five components required of effective teaching: planning instruction and assessment; instructing and engaging students in learning; assessing student learning; analysis of teaching effectiveness; and academic language development (SCALE, 2013b).

The edTPA was collaboratively designed by teacher educators and teachers under the coordination of Stanford University and with support from the American Association for Colleges of Teacher Education (AACTE) (Sato, 2014). Pearson Education provides edTPA materials, training, and virtual support in order to administer the edTPA on a national scale (SCALE, 2013b). In Minnesota, each college or university’s director of student teaching coordinates completion of the edTPA each semester.
**Type of response categories.** According to SCALE (2013b), teacher candidates submit teaching artifacts and reflective commentaries to demonstrate their proficiency as pre-service teachers, and to justify the instructional decisions they have made during the learning segment:

- Artifacts represent authentic work completed by the teacher candidate and his/her students (e.g., lesson plans, copies of instructional and assessment materials, unedited video recording(s) of the candidate’s teaching, and student work samples).

- Commentaries require candidates to explain the artifacts, justify the rationale behind the choice of artifact or instructional decision, and analyze what he/she has learned about students’ learning and the effectiveness of his/her teaching practice. (p. 9)

**Appropriateness of instruments to the study.** Establishing validity and reliability are essential elements of testing and utilizing data collection instruments (Roberts, 2010). “Evaluating the validity and reliability of edTPA as an instrument to measure teacher candidates’ readiness to teach has been a continuous part of its rigorous, multi-year development process” (SCALE, 2013a, p. 17).

According to SCALE (2013b), educational stakeholders developed the edTPA “within a technical framework of psychometric practice and principles guided by The Standards for Educational and Psychological Testing (AERA, APA, & NCME, 1999)” (p. 17). Beginning with the Performance Assessment for California Teachers (PACT), development and refinement of the nationally available edTPA produced evidence to establish the assessments’ validity, reliability, and utilization for high-stakes accountability in the field of education. For example, the edTPA has been implemented nationally, and each state may choose to utilize the edTPA for teacher licensure and teacher preparation programs approval and accreditation (SCALE, 2013b). “This work has proceeded under the guidance of experts in psychometric practices and procedures. In addition, participating states have subjected the analyses to further review by their
own technical advisors. Advisors have held the process to a high standard” (SCALE, 2013a, p. 17).

**Content validity.** In order establish content validity, researchers at SCALE (2013b) gathered “feedback from educators to rate the importance, alignment, and representativeness of the knowledge and skills required for each rubric, and of the rubric itself, in relation to national pedagogical and content-specific standards” (p. 20). Establishing content validity of the edTPA provides essential evidence of relatedness to the knowledge, skills, and abilities necessary for teaching (SCALE, 2013b).

Content validation activities asked reviewers to rate the knowledge and skills of the edTPA on a 5-point scale. Content validation results showed that all three tasks and components rated above 3 on the 5-point scale, revealing tasks as having “importance” to “very great importance” (SCALE, 2013b). Relating to the InTASC standards, content validation results indicated that all tasks and edTPA components align to the standards “well” to “very well” (SCALE, 2013b). Finally, content validity results determined that the edTPA rubrics represent an assessment of knowledge and skills required of the act of teaching. According to SCALE (2013b), “To further support content validity, an additional round of activities was conducted to provide additional confirmation of the importance, alignment, and representativeness of the edTPA tasks” (p. 20).

**Construct validity.** Researchers at SCALE (2013b) used factor analysis of the composite edTPA score to deduce a candidate’s readiness to teach. An exploratory factor analysis results in a set of estimated factor loadings, which can range from -1.0 to +1.0 (Kuzma & Bohnenblust, 2001). Researchers conducted factor analyses of the 15 edTPA rubrics organized by task (i.e. Task 1: Planning; Task 2: Instruction; and Task 3: Assessment). Results revealed that the five
rubrics contained within the same task—i.e. Task 1: Planning; Task 2: Instruction; and Task 3: Assessment—were closely related, and that candidate performance was also positively related within the tasks (SCALE, 2013b). Thus, SCALE (2013b) concluded:

The data suggest that all factor loadings are positive and of moderate to large magnitude. These results support the use of a single total score. The second factor analysis demonstrated that the hypothesized task structure of the edTPA is supported by the patterns of candidates’ scores. (p. 23)

**Reliability.** As a nationally available performance assessment used for teacher licensure and teacher preparation program improvement, researchers at SCALE (2013b) sought statistically viable evidence to establish the consistency of edTPA scores. Researchers conducted two types of analyses during the 2013 national edTPA field test. “First, the agreement rates between different scorers evaluating the same candidate’s submission were analyzed. Second, the overall variability in a candidate’s scores due to chance measurement error was evaluated using a Cohen Kappa statistical procedure” (SCALE, 2013b, p. 23).

In order to analyze the consistency of edTPA scores, “approximately 10% of all edTPA submissions are randomly selected to be scored by a second, independent scorer” (SCALE, 2013b, p. 23). Researchers predicted that if the edTPA is a reliable assessment, candidate scores would be scored the same or have small differences among the two scorers within the random sample (SCALE, 2013b). Hoover (2012) reported an internal consistency rate (Cronbach’s alpha) of the edTPA at .93.

**Inter-rater reliability procedures.** SCALE (2013b) conducted several reliability analysis in order to determine agreement rates among edTPA evaluators. Appendix D provides a detailed analysis of both adjacent agreement rate and kappa – n statistics (SCALE, 2013b). Because of the complexity of the edTPA, SCALE employs adjacent agreement in order to establish inter-
rater reliability. According to SCALE (2013b), “Adjacent agreement refers to the proportion of cases in which two independent scorers assign either the exact same score or a score within 1 point of each other” (p. 23).

Inter-related reliability is affected by chance, as opposed adjacent agreement as a result of the training of edTPA evaluators. Researchers at SCALE (2013b) employed the “kappa \(-n\)” statistic in order to account for agreement by chance in the edTPA evaluation process. According to SCALE (2013b), “The values reported here, which are generally greater than 0.8 (minimum of 0.65 and maximum of 0.921) suggest scorers are nearly always within 1 point” (p. 23). The table in Appendix D illustrates the inter-rater reliabilities of both adjacent agreement and agreement by chance. To add, SCALE (2013b) reported:

The average adjacent agreement rate was .92 and the average kappa \(-n\) was .83. These reliability levels are relatively high, and are comparable to those found for well-established performance assessments. Assessing the degree of agreement among edTPA scorers, research indicates edTPA reliabilities ranging from .83 to .92. (p. 24)

**Administration and scoring rubrics.** Trained evaluators score a teacher candidate’s full edTPA submission. The edTPA is a subject-specific assessment, and trained evaluators possess extensive content understanding (SCALE, 2013b). Candidates are evaluated on five key components of teaching quality (SCALE 2013b):

1. Planning (contained within Task 1 of the edTPA)
2. Instruction (contained within Task 2 of the edTPA)
3. Assessment (contained within Task 3 of the edTPA)
4. Analysis of Teaching (contained within all edTPA Tasks)
5. Academic Language (contained within all edTPA Tasks)
A single edTPA evaluator grades a teacher candidate’s entire edTPA and possesses the grade level and subject area expertise (SCALE, 2013b). During the review process, the evaluator analyzes the evidence in which the teacher candidate provides in order to demonstrate their knowledge, skills, and application of edTPA elements (SCALE, 2013b). The edTPA is comprised of 15 different elements—five rubrics for each of the three tasks (SCALE, 2013b). A candidate can achieve a score ranging from a minimum of 15 points to a maximum of 75 points. Evaluators utilize graduated rubrics to evaluate 15 different aspects of teaching quality.

Individual rubric grades range from Level 1 to Level 5 (SCALE, 2013b, p. 12):

- Level 1: Represents the knowledge and skill level of a struggling candidate who is not ready to teach.
- Level 2: Represents the knowledge and skills of a candidate who is possibly ready to teach.
- Level 3: Represents the knowledge and skills of a candidate who is ready to teach.
- Level 4: Represents a candidate with a solid foundation of knowledge and skills for a beginning teacher.
- Level 5: Represents the advanced skills and abilities of a candidate very well qualified and ready to teach.

The development of rubrics by education stakeholders is grounded in research and assesses key elements of teaching quality (SCALE, 2013b). According to researchers at SCALE (2013b), the edTPA evaluates:

... how teachers plan to support learning goals and student needs, create a positive learning environment, engage students in ways that deepen their learning, create meaningful tasks and provide feedback on student work, support student learning of language, and analyze their teaching in relation to students’ learning and make further plans to facilitate their progress. The criteria emphasize how teachers support the learning of all students by understanding their students’ prior knowledge, experiences, and cultural contexts, and by teaching with an eye toward equity. (p. 12)
Data Collection Procedures

College of Education staff administers the edTPA two times each academic year: fall semester and spring semester. The university has implemented timelines and procedures for teacher candidates for successful technical completion of the edTPA. As part of the procedure, teacher candidates purchase a subscription to a learning management system called TaskStream that houses the necessary materials to complete the performance assessment. TaskStream is then linked to Pearson Education, and Pearson Education sends the completed edTPA to trained scorers for evaluation. Pearson Education and SCALE then send the results of the scored edTPA back to the university and to the individual teacher candidate.

Data Analysis

edTPA data are reported to the selected university in aggregate as a total assessment score, disaggregated by edTPA task (Task 1: Planning; Task 2: Instruction; and Task 3: Assessment), and scores by individual rubric—15 total rubrics (five rubrics per task).

Teacher candidate edTPA performance data will be displayed as numerical scores. The numerical scores of participants will be matched with responses on the CMSS by their supervisors to run regression and correlation statistical analyses.

The regression analysis will be implemented to answer research questions one through four. The correlation analysis will be employed to answer research question number five. The regression analysis will be used to explore if there is a predictive relationship between edTPA scores and supervisor responses on the CMSS. The correlation analysis will be used to explore if the two instruments are generally related as a measure of teaching quality.
**CMSS.** The Network for Excellence in Teaching (NExT), a partnership of 14 Institutions of Higher Education (IHEs) and the Bush Foundation, aims to transform how university-based teacher education programs prepare new, effective teachers in the Midwest.

The NExT institutions have collaborated to develop and administer a set of four common surveys to measure the IHE’s progress toward this goal. Teacher candidates and program completers at each of the IHEs complete three surveys upon entry into the teacher education programs; at exit; and one year after graduation (known as the Transition to Teaching Survey or TTS). Supervisors of NExT program completers working in the teaching field complete the Supervisor Survey (CMSS) after the graduate’s first year of teaching.

The CMSS seeks to understand perceptions of those who supervise first-year teachers of the novices’ readiness for the teaching profession. The survey evaluates supervisor’s perceptions of the quality of graduates’ instructional practices, abilities to work with diverse learners, abilities to establish positive classroom environments, and levels of professionalism. The survey is administered to direct supervisors of teacher education graduates employed in the schools approximately one year after the teachers completed their preparation programs.

In 2013, NExT IHEs were responsible for providing contact information for their 2011-2012 graduates and their graduates’ supervisors to an external research agency, Hezel Associates, which administered the CMSS. In 2014, many IHE’s agreed to self-administer the CMSS in an effort to increase response rate. Hezel Associates continued to administer the survey for other IHE’s.

Administration sites—i.e. individual colleges and universities in NExT—conducted many of the same survey administration protocols. However, each IHE used its discretion to determine the most effective way to administer the survey and increase response rates. The selected
university was one of the IHE’s that agreed to separately administer the survey and sent an electronic version of the survey to eligible supervisors via email and provided several reminders. Among the university’s respondents, all 85 reported that they were prompted to complete the CMSS by an email message. Intensive efforts by the university to collect accurate contact information and promote survey completion led to a greater number of survey responses and more useful data sets.

**Response rate.** The 2014 CMSS response rate for the university was 31% (85 out of 271). In comparison, the NExT aggregate response rate was 16%. This response rate is calculated by using the number of 2012-2013 teacher education program computers as the denominator and the number of program completer’s supervisors who responded to the survey as the numerator. The number of 2012-2013 program completers was reported by IHEs when the institutions submitted their Exit Survey data. By comparison, the university’s response rate in 2013 was 5%—very low in order to draw concrete conclusions.

**Reliability and validity.** Hezel Associates, a third-party research agency, conducted validity and reliability testing on the CMSS. Results of validity and reliability testing indicated the need to modify questions between survey administrations of the 2012-2013 and 2013-2014 academic years. I used the most recent CMSS for the study to ensure survey item continuity and consistency of the data.

**Data Collection Procedures**

Data were collected using the Qualtrics survey application sent via email to respondents. Names and directory information of supervisors was self-reported by new teachers in order for the university to send the CMSS via email. Follow-up emails were sent to the supervisors from the Dean’s office if supervisors had not responded after the first attempt.
Data were collected from May to July each academic year. Data were housed within the Qualtrics survey application and were accessible by the College of Education’s data and assessment managers. Raw data is accessible upon written request with a data request form. Data and assessment managers analyzed the data and wrote descriptive reports in order for faculty to improve curriculum and programming.

**Data Analysis**

Common Metrics Supervisor Survey response data will be matched to the supervisor’s new teacher and the new teacher’s edTPA scores. Through comparative analysis, CMSS data will be analyzed in relation to edTPA scores in aggregate and disaggregated by edTPA “Task.” Results of correlation and regression statistical analyses will be reported with descriptive statistics and displayed via tables.

In order to answer Research Questions 1-5, I will employ regression statistical analyses using data from the CMSS and edTPA after matching identifying information.

**Limitations**

Limitations of the method include using the CMSS and edTPA for an individual program completer. This limitation—having complete CMSS and edTPA scores—will narrow the focus in answering the research questions.

**Summary**

The Common Metrics Supervisor Survey is an instrument that collects perceptions of new teacher’s teaching quality by their supervisor. The edTPA performance assessment provides aggregate and disaggregated numerical scores on a teacher candidates teaching quality during student teaching. Data from the CMSS and edTPA will be comparatively analyzed using regression and correlation statistical analyses in order to answer the research questions.
Chapter IV: Results

Introduction

Multiple researchers concluded that the most impactful school-based factor that affects student learning is the quality of teaching in the classroom (Goldhaber, 2007; McCaffrey et al., 2003; Rivkin et al., 2000; Rockoff, 2004; Rowan et al., 2002; Wright et al., 1997). As a profession, effective metrics assessing teaching quality must be implemented to identify, justify, and improve teaching and learning in schools (Darling-Hammond, 2014). The edTPA (Teacher Performance Assessment) and Common Metrics Supervisor Survey (CMSS) are metrics designed to assess such teaching quality.

The edTPA and CMSS are available data sources utilized by the small public university in Minnesota to improve teacher preparation programming at the selected university. Figure 1 displays four thematic categories in which the 10 InTASC Model Core Teaching Standards are grouped.

![InTASC Model Core Teaching Standards Categories](image)

*Figure 1. InTASC Model Core Teaching Standards Categories*
As displayed in Table 2, both edTPA and CMSS instruments are research-based and aligned to the national InTASC professional teaching standards (Hezel Associates, 2013; SCALE, 2014).

Table 2

**edTPA and CMSS Alignment to InTASC Core Teaching Standards**

<table>
<thead>
<tr>
<th>Data Source</th>
<th>InTASC Model Core Teaching Standards</th>
<th>InTASC Model Core Teaching Standards</th>
<th>InTASC Model Core Teaching Standards</th>
<th>InTASC Model Core Teaching Standards</th>
</tr>
</thead>
<tbody>
<tr>
<td>edTPA Rubrics</td>
<td>Category 1: The Learner and Learning (Standards 1-3)</td>
<td>Category 2: Content (Standards 4-5)</td>
<td>Category 3: Instructional Practice (Standards 6-8)</td>
<td>Category 4: Professional Responsibility (Standards 9-10)</td>
</tr>
<tr>
<td></td>
<td>1, 2, 3, 4, 5, 6, 7, 8, 9, 14</td>
<td>1, 2, 3, 4, 7, 8, 9, 14</td>
<td>10, 15</td>
<td></td>
</tr>
<tr>
<td>Common Metrics Supervisor Survey Questions</td>
<td>B1a, B1b, B1c, B3b</td>
<td>B1b, B1o, B1r, B1e, B1f, B1g, B1h, B1i, B1j, B1k, B1l, B1m, B2j</td>
<td>B4a, B4b, B4c, B4d, B4e, B4f</td>
<td></td>
</tr>
<tr>
<td></td>
<td>B1d, B1n, B1p, B1q, B1s, B2a, B2b, B2c, B2d, B2e, B2f, B2g, B2h, B2i, B3a, B3c, B3d, B3e, B3f, B3g, B3h, B3i</td>
<td>B1a, B1b, B1c, B1d, B1e, B1f, B1g, B1h, B1i, B1j, B1k, B1l, B1m, B1n, B1o, B1p, B1q, B1r, B1s, B2a, B2b, B2c, B2d, B2e, B2f, B2g, B2h, B2i, B3a, B3c, B3d, B3e, B3f, B3g, B3h, B3i</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2 displays the alignment of the edTPA Rubrics and CMSS questions to the InTASC Core Teaching Standards. SCALE (2014) researchers produced the edTPA alignment analysis, while the Common Metrics group completed the CMSS alignment analysis (Hezel Associates, 2012). The researcher combined the individual analyses into Table 2. As shown, the edTPA contains 15 total rubrics. The CMSS contains 45 total items. InTASC Category 1: The Learner and Learning (Standards 1-3) aligns with 10 edTPA rubrics: 1, 2, 3, 4, 5, 6, 7, 8, 9, and 14. Related, 23 CMSS items align to InTASC Category 1: B1d, B1n, B1p, B1q, B1s, B1t, B2a, B2b, B2c, B2d, B2e, B2f, B2g, B2h, B2i, B3a, B3c, B3d, B3e, B3f, B3g, B3h, and B3i.

InTASC Category 2: Instructional Practice (Standards 4-5) aligns with 13 edTPA rubrics: 1, 2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 13, and 15. Three CMSS items align to InTASC Category 2: B1a, B1c, and item B3b. InTASC Category 3: The Learner and Learning (Standards 6-8) aligns with
10 edTPA rubrics: 1, 2, 3, 4, 5, 6, 7, 8, 9, and 14. Thirteen CMSS items align to InTASC Category 3: B1b, B1o, B1r, B1e, B1f, B1g, B1h, B1i, B1j, B1k, B1l, B1m, and B2j. Finally, InTASC Category 4: Professional Responsibility (Standards 9-10) aligns with two edTPA rubrics: 10 and 15. Six CMSS items align to InTASC Category 4: B4a, B4b, B4c, B4d, B4e, and item B4f.

**Research Problem/Purpose**

According to Darling-Hammond (2010), performance assessments in the pre-service stage of preparing educators (such as the edTPA) may provide answers to questions regarding predictive metrics of teaching quality. Specific to evaluating teaching quality at the completion of the first year of teaching, the CMSS is administered to determine a supervisor’s perceptions of a new teacher’s quality of instructional practice, educating diverse learners, the classroom environment, and professional responsibilities. The problem is that the instruments—edTPA and CMSS—are used separately by participating IHE’s.

The purpose of the study was to examine the relationship between edTPA scores and perceived teaching quality after the first year of teaching by a beginning teacher’s supervisor on the CMSS. The edTPA is a valid and reliable assessment of teaching quality (SCALE, 2013b). A third party contractor conducted a validity study (Hezel Associates, 2012) and a reliability study (Hezel Associates, 2013) of the Common Metrics Supervisor Survey.

**Research Questions**

Drawing from the literature, the researcher developed five research questions to guide the study. Due to the importance of aligning outcomes to the professional teaching standards, the researcher organized the research questions to align with the national InTASC Core Teaching Standards (CCSSO, 2011):
1. To what extent do overall edTPA scores predict teaching quality at the end of the first year as determined by supervisor responses on the Common Metrics Supervisor Survey?

2. To what extent do edTPA scores predict CMSS responses related to “InTASC Category 1: The Learner and Learning?”

3. To what extent do edTPA scores predict CMSS responses related to “InTASC Category 2: Content?”

4. To what extent do edTPA scores predict CMSS responses related to “InTASC Category 3: Instructional Practice?”

5. To what extent do edTPA scores predict CMSS responses related to “InTASC Category 4: Professional Responsibility?”

**Method**

For the study, College of Education staff acquired teacher candidate edTPA scores from Pearson Education for the first data set. For the second data set, supervisor perceptions of a new teacher’s quality were obtained using the Common Metrics Supervisor Survey (CMSS) instrument’s five-point Likert scale (i.e., strongly disagree = 1, disagree = 2, agree = 3, strongly agree = 4, not able to respond = 0). Incomplete responses were extrapolated using a linear trend of the subject’s other responses (Kuzma & Bohnenblust, 2001). The criterion-referenced predictive validity study was completed using linear regression statistical analyses in order to answer Research Questions 1-5.

Administration of the edTPA to teacher candidates was accomplished during their student teaching experiences, which typically occurred during the final year of their preparation programs. Administration of the CMSS to supervisors occurred at the completion of the new teacher’s first year through Qualtrics survey software. R statistical programming language software was used to calculate mean scores for each of the 45 CMSS survey items. Discussed in more detail below, data were cleaned prior to entering into R software. Official edTPA data used
in the study were all available edTPA records from 2013 through spring 2016 submissions. CMSS data used in the study from the selected university were all available records from 2014 through 2016.

Teacher candidates’ first and last names were utilized to match the their official edTPA score reports and the submitted Common Metrics Supervisor Surveys. Teacher candidates’ TechID were used to verify the match between the two data sets. After matching, candidates’ first and last names, as well as TechID, were removed from the data set to insure candidates’ anonymity. To be retained for analysis, cases needed to:

- Have valid edTPA data (must not contain condition codes—discussed below).
- Have valid survey data (minimum of 30 item responses; maximum of 10 responses of “Does not apply” which was coded as “0”).

In order for edTPA data to be valid and included in the study, all 15 edTPA rubrics were only to contain numerical values (1-5 points) and were not to contain place-holding letters called “condition codes.” Condition codes were scored as zeros, in order to describe a large error in the submitted edTPA assessment. An evaluator issued a condition code to a candidate’s rubric as a result of the following criteria (Pearson Education, 2016):

- Code A: Rubrics for Planning Task are unable to be scored; fewer than three lesson plans; hyperlinks instead of actual artifacts.
- Code B: Video Technical Issues; must meet technical specs and must review in Pearson to make sure it can be played.
- Code C: Audio Technical Issues; unintelligible to scorers; no audio is heard.
• Code D: Insufficient information to score; too little commentary; commentary not within the page limits.

• Code E: Incorrect or missing, blank, or otherwise inaccessible file.

• Code F: Video is edited; videos must be continuous and unedited.

• Code G: Materials unrelated to Handbook or fail to conform to handbook instructions; content must match handbook; learning objectives not in right area; no central focus.

• Code H: English translation requirement not met; if part of the edTPA is not in English, required translation/transcript is required.

In sum, a teacher candidate may have made an error with any of the above items on the final submission of their edTPA. In these instances, edTPA evaluators issued condition codes in order to explain the error to the candidate and the IHE. For the study, any edTPA containing at least one of the aforementioned condition codes was excluded from the study.

The results of the study are reported descriptively according to the five research questions that were developed in order to reveal the relationship between the edTPA and CMSS. Also, a cumulative score for this section of the survey was obtained to compare the mean difference between the edTPA and the CMSS. The level of significance to which the study was held was <.05. Items on the edTPA instrument were empirically justified (SCALE, 2013b). CMSS data were subjected to confirmatory factor analysis (Hezel Associates, 2013), which defined a relationship between the items assigned to each factor. Cronbach alpha scores greater than .70 were considered acceptable for internal reliability of each factor (Peterson, 1994). R statistical
programming language also generated other descriptive statistics (i.e., percentages, mean scores, and standard deviations) that were used for item analysis (R Core Team, 2016).

EdTPA submissions for the selected university occurred during the fall and spring semesters of each academic year. Teacher candidates at the selected university first completed the EdTPA during the fall semester of 2013. The study examined EdTPA submissions for the 2013, 2014, 2015, and 2016 academic years. The total number of cases within the EdTPA data set prior to cleaning equaled 555 (n=555), respectively. To clean the EdTPA data set, any case containing at least one condition code was removed from the final data set. The EdTPA data set changed from n=555 to n=508 after applying this limiting rule.

According to the Common Metrics Supervisor Survey technical manual, College of Education staff members were to administer the CMSS between April and June of an academic year. The total number of CMSS cases totaled 257 (n=257) over the three years of administration (2014, 2015, and 2016). CMSS response rates for each year resulted in 2014 (n=79); 2015 (n=91); 2016 (n=87). Contained within the raw CMSS data were “998,” “999,” and “0” placeholders. Placeholders with a “998” revealed that a question was viewed and skipped. Placeholders with a “999” indicate that surveys had been opened and abandoned, while “0” placeholders indicated that the respondents answered “does not apply.” For instance, a principal may have skipped forward in a survey if they did not have knowledge of the new teacher’s engagement in professional development activities. In this situation, the value placed for the survey item would be “998.” In the case in which a supervisor opened a survey and, subsequently, closed the survey (and did not complete the remaining items), the values placed for each unanswered item would be “999.” Finally, in some instances a supervisor may have felt that
he or she did not have sufficient information to answer the question. In these cases, respondents may have elected to select “does not apply,” thus creating a value of “0” for survey items.

When CMSS data were examined, four cases had less than 30 items completed of 45 total survey items. Four cases were removed from the dataset as a result of containing 10 or more “zero” values of the 45 items on the survey. After applying these two rules (<30 survey items completed; >10 “zero” values), the final total of CMSS cases equaled 249 (n=249).

Using TechID to match the two data sets, the total number of candidates who had both edTPA and CMSS data equaled 120 (n=120). After cleaning the edTPA dataset, 14 cases were removed for containing any of the applied limiting rules, resulting in a final number of 106 (n=106). The matched “cleaned” data were entered into a data analysis software program and regression analyses were conducted.

**Results for Research Questions**

This chapter reports the findings of the study. The findings described below are based upon a set of research questions that were developed by the researcher and emerged from the literature. The researcher developed five research questions in order to examine the relationship between edTPA and CMSS and teaching quality. The researcher conducted regression analyses on existing data to answer Research Questions 1-5. Two existing data sets were compared in the study: edTPA and CMSS.

For edTPA data, the Pearson Education group reported official edTPA scores to College of Education (COE) staff in the following format:

1. An overall score on the edTPA (Ex: 44 out of 75 possible points);
2. Scores reported per task (Ex: Task 1 Planning = 13 out of 25 possible points; Task 2 Instruction = 14 out of 25 points; Task 3 Assessment = 12 out of 25 points);
3. Scores reported per rubric (Ex: Rubric #3 = 3 out of 5 points; Rubric #13 = 2 out of 5 points).

College of Education staff, including the researcher, collected CMSS data using Qualtrics survey software and managed data internally. Export of data to .csv spreadsheets allowed COE staff to conduct analyses of CMSS responses. The data were analyzed and findings organized according to each of the following research questions:

1. To what extent do overall edTPA scores predict teaching quality at the end of the first year as determined by supervisor responses on the Common Metrics Supervisor Survey?

2. To what extent do edTPA scores predict CMSS responses related to “InTASC Category 1: The Learner and Learning?”

3. To what extent do edTPA scores predict CMSS responses related to “InTASC Category 2: Content?”

4. To what extent do edTPA scores predict CMSS responses related to “InTASC Category 3: Instructional Practice?”

5. To what extent do edTPA scores predict CMSS responses related to “InTASC Category 4: Professional Responsibility?”

A third-party consultant completed the analysis of data using the R statistical programming language (R Core Team, 2016). The researcher selected the third-party consultant, a former colleague, due to his experience administering the CMSS for the 2013-2015 administrations at the selected university. The third-party consultant earned a Master’s Degree in Social Psychology and has extensive experience analyzing edTPA and Common Metrics data for the selected university in the study.

The first data set was based on all available edTPA submissions (n=555). The second data set was based on all available Common Metrics Supervisor Surveys (n=257). After
matching the two data sources together using a participant-level unique identifier, and cleaning
the dataset, there were n=106 complete records available for analysis.

EdTPA data are reported in aggregate and disaggregated by “Task” in Table 3 below. Table 3 provides a base in which the researcher was able to ascertain range of scores, quartile values, median, and mean scores in relation to the research questions.

Table 3 reports basic descriptive statistics of teacher candidate overall edTPA scores at
the selected university (n=106). Row one describes statistics for the full edTPA assessment (scores could range from 15 to 75). Rows two, three, and four describe scores by “Task” (Task 1: Planning; Task 2: Instruction; Task 3: Assessment). Each “Task” contains five rubrics, and scores can range from 1-5 points per rubric.

Table 3

| edTPA Descriptive Statistics Based on Pearson Education Reports (n=106) |
|-----------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|
|                             | Min  | 1st Quartile | Median | Mean  | 3rd Quartile | Max  |
| edTPA                      | 22.00 | 35.00   | 42.00 | 40.71 | 46.00   | 59.00 |
| Task 1                     | 7.00  | 12.25 | 14.25 | 14.26 | 16.00 | 20.00 |
| Task 2                     | 7.00  | 12.00  | 14.00 | 13.86 | 15.38 | 19.00 |
| Task 3                     | 5.00  | 11.00  | 13.00 | 12.59 | 14.88 | 20.00 |

The minimum overall edTPA score was 22.00 points; the median score was 42.00 points; and the maximum score was 59.00 points. The overall mean score for the n = 106 cases equated 40.71 points. Examined by “task,” Task 1: Planning contained the highest mean of 14.26 points (out of 25.00 possible points). Next, Task 2: Instruction contained the second highest mean of 13.86 out of 25.00 possible points. Finally, Task 3: Assessment exhibited a mean of 12.59 points out of a possible 25.00 points.
Table 4 reports basic descriptive statistics of teacher candidate edTPA scores grouped by alignment to the InTASC Core Standard Categories. Values are displayed in points earned on evaluated rubrics, and rubric scores can range between 1.00-5.00 points.

Table 4

<table>
<thead>
<tr>
<th>edTPA Category</th>
<th>Min</th>
<th>1st Quartile</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Quartile</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>edTPA Category 1: The Learner</td>
<td>15.00</td>
<td>24.62</td>
<td>29.00</td>
<td>27.97</td>
<td>31.00</td>
<td>39.00</td>
</tr>
<tr>
<td>edTPA Category 2: Content</td>
<td>11.00</td>
<td>19.12</td>
<td>23.00</td>
<td>22.23</td>
<td>25.00</td>
<td>32.00</td>
</tr>
<tr>
<td>edTPA Category 3: Instructional Practice</td>
<td>20.00</td>
<td>31.00</td>
<td>37.00</td>
<td>35.78</td>
<td>40.00</td>
<td>51.00</td>
</tr>
<tr>
<td>edTPA Category 4: Professional Responsibility</td>
<td>2.00</td>
<td>4.00</td>
<td>5.00</td>
<td>5.05</td>
<td>6.00</td>
<td>8.00</td>
</tr>
</tbody>
</table>

Note: Total possible points on an individual rubric is 5.00 points.

Table 4 displays the disaggregated edTPA data in four InTASC Core Teaching Standards Categories. InTASC Category 1: The Learner and Learning edTPA scores revealed a minimum score of 15.00 points, median score of 29.00 points, and maximum score of 39.00 points. The Category 1 total mean score equated to 27.97 points. Category 2: Content edTPA scores ranged from a minimum of 11.00 to maximum of 32.00 points, with a median score of 23.00 and mean of 22.23 points. Category 3: Instructional Practice edTPA scores ranged from 20.00 to 51.00 points, with a median of 37.00 points and mean of 35.78 points. Finally, Category 4: Professional Responsibility edTPA scores ranged from 2.00 to 8.00 points, with a median of 5.00 points and mean of 5.05 points.

Table 5 describes basic descriptive statistics of supervisor perceptions as captured by the Common Metrics Supervisor Survey. CMSS items were grouped by alignment to the InTASC
Core Standard Categories. CMSS Likert scale responses contain values ranging between 0 and 4 (not able to respond = 0, strongly disagree = 1, disagree = 2, agree = 3, strongly agree = 4).

Table 5

CMSS Descriptive Statistics Based on InTASC Standard Categories (n=106)

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>1	extsuperscript{st} Quartile</th>
<th>Median</th>
<th>Mean</th>
<th>3	extsuperscript{rd} Quartile</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMSS Average</td>
<td>2.42</td>
<td>3.28</td>
<td>3.63</td>
<td>3.55</td>
<td>3.91</td>
<td>4.00</td>
</tr>
<tr>
<td>CMSS Category 1:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The Learner and</td>
<td>2.04</td>
<td>3.13</td>
<td>3.61</td>
<td>3.51</td>
<td>3.91</td>
<td>4.00</td>
</tr>
<tr>
<td>Learning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CMSS Category 2:</td>
<td>2.00</td>
<td>3.67</td>
<td>4.00</td>
<td>3.69</td>
<td>4.00</td>
<td>4.00</td>
</tr>
<tr>
<td>Content</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CMSS Category 3:</td>
<td>2.27</td>
<td>3.26</td>
<td>3.56</td>
<td>3.53</td>
<td>4.00</td>
<td>4.00</td>
</tr>
<tr>
<td>Instructional</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Practice</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CMSS Category 4:</td>
<td>2.00</td>
<td>3.50</td>
<td>3.67</td>
<td>3.65</td>
<td>4.00</td>
<td>4.00</td>
</tr>
<tr>
<td>Professional</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Responsibility</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The number of CMSS survey items within each of the four InTASC Core Teaching Standards Categories varied by category. Category 1: The Learner and Learning comprised 23 survey items. Category 2: Content contained three survey items. Category 3: Instructional Practice included 13 items. Finally, Category 4: Professional Responsibility contained six survey items. Thus, the degree to which descriptive statistics vary was based upon the number of survey items in each category.

Overall CMSS values ranged from a minimum of 2.42 to a maximum value of 4.00. The overall median was 3.63 with a mean of 3.55. Category 1 values ranged from a minimum of 2.04 to a maximum of 4.00, with a median score of 3.61 and a mean of 3.51. Category 2 values ranged from a minimum of 2.00 to a maximum of 4.00, with a median score of 4.00 and a mean of 3.69. Category 3 values ranged from a minimum of 2.27 to a maximum of 4.00, with a median score of 3.56 and a mean of 3.53. Finally, Category 4 values ranged from a minimum of 2.00 to a
maximum of 4.00, with a median score of 3.67 and a mean of 3.65. To reiterate, the overall CMSS mean and all four InTASC category mean values fell between “agree” and “strongly agree.”

The first research question was intended to examine the whether or not overall scores on the edTPA predicted supervisor perception outcomes on the CMSS.

**Research Question 1.** To what extent do overall edTPA scores predict first-year teaching quality as determined by supervisor responses on the Common Metrics Supervisor Survey?

Several statistical tests were conducted to answer five separate research questions regarding the relationship between edTPA scores and Common Metrics Supervisor Survey (CMSS) responses. Analysis of the overall assessment tools provided an examination of all essential components of teaching as a profession.

Four simple linear regressions were computed using R statistical programming language (using the “lm” command in the base {stats} package for the R statistical program) and one single Pearson correlation coefficient was computed (using the “cor.test” command in the base {stats} package in R) (Grothendieck, 2014).

To answer Research Question 1, the researcher conducted a linear regression statistical test to determine the effect of the relationship between overall edTPA scores and overall CMSS results.

Table 6 reports the results of a linear regression analysis of the overall edTPA scores’ relationship with the overall CMSS results.
A linear regression was computed using edTPA overall scores as the single predictor variable and CMSS overall average scores as the dependent variable. A statistically significant ($p = 0.04$) model was revealed: edTPA was found to be a significant predictor of CMSS average scores. Higher edTPA scores predicted higher CMSS responses. Based on the model, every 10-point increase on the edTPA is associated with a 0.12 point increase on the CMSS average (on a 4-point scale). Furthermore, edTPA scores accounted for approximately 3% of the variance in CMSS scores. Although this model is significant, the size of this effect was small. Figure 2 below shows these two variables on a scatter plot with the best-fit regression line (in red) superimposed.

### Table 6

*Linear Regression: Use Overall edTPA Score to Predict CMSS Overall Average Score (n=106)*

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t value</th>
<th>p value</th>
<th>Adj. R2</th>
</tr>
</thead>
<tbody>
<tr>
<td>edTPA Score</td>
<td>0.012</td>
<td>0.006</td>
<td>2.069</td>
<td>0.041 *</td>
<td>0.030</td>
</tr>
</tbody>
</table>
Figure 2. edTPA Scores and CMSS Overall Average Score (n=106)

Figure 2 illustrates the distribution of overall edTPA scores on the x axis, and the y axis displays the values of overall responses on the CMSS. Each point on the plot represents a matched case. The plotted data of Figure 2 illustrates a concentration of relatively high CMSS values between 3.5 and 4.0. The best-fit regression line (in red) demonstrates an upward trend, or significant positive relationship, between the two variables of the linear regression ($p = .04$). The best-fit regression line provides a visual representation of the significant positive correlation among the variables analyzed to answer the first research question. Although there were a number of “outlier” points in the data, the concentration of matched cases fell within the best-fit
regression line. Thus, Figure 2 illustrates findings that the higher the overall edTPA score, the higher the overall CMSS response score.

Four additional linear regressions were computed on the same dataset using specific survey items and edTPA rubrics that were hypothesized to be related based on the InTASC standards.

**Research Question 2.** To what extent do edTPA scores predict CMSS responses related to “InTASC Category 1: The Learning and Learning?”

The Interstate Teacher Assessment and Support Consortium (InTASC) of the Council of Chief State School Officers (CCSSO) developed the national teacher education standards (CCSSO, 2013). The education field has accepted these standards as desired professional goals to achieve in order to improve teaching and, subsequently, student learning (CCSSO, 2013).

To address the second research question, the researcher matched the two data sources together using a participant-level unique identifier and cleaned the data, resulting in n=106 complete records for analysis. The researcher then grouped edTPA rubric scores and CMSS item values into InTASC Core Teaching Standards-aligned categories. Table 6 reports the results of a linear regression analysis of edTPA scores’ relationship with CMSS results related to InTASC Core Teaching “Category 1: The Learner and Learning” in relation to the research questions.

Table 7 provides the descriptive statistics of estimate, standard error, t-value, p-value, and adjusted R².
A linear regression was computed using edTPA Category 1 score as the single predictor variable and CMSS Category 1 average scores as the dependent variable. A statistically significant (p = 0.036) model was revealed: Category 1 edTPA score was found to be a significant predictor of Category 1 CMSS scores. Higher edTPA scores predicted higher CMSS responses. Based on the model, every ten-point increase on the Category 1 edTPA scores was associated with a 0.21 point increase on the Category 1 CMSS average. Furthermore, edTPA scores accounted for approximately 3% of the variance in CMSS scores. Although this model was significant, the size of this effect was small.

Figure 3 shows a scatter plot with associated best-fit line for the regression model tested.
Figure 3. edTPA Category 1: The Learner and Learning scores and CMSS Category 1: The Learner and Learning scores

Figure 3 shows the distribution of Category 1: The Learner and Learning edTPA scores on the x axis, and the y axis displays the values of Category 1: The Learner and Learning CMSS responses. The best-fit regression line (in red) demonstrated an upward trend, or significant positive relationship, between the two variables of the linear regression (p = .04). The plotted data of Figure 3 demonstrated a concentration of relatively high edTPA scores between 27 and 35. The best-fit regression line (in red) demonstrated an upward trend, or significant positive relationship, between the two variables of the linear regression (p = .04). The best-fit regression line provided a visual representation of the significant positive correlation among the variables.
analyzed to answer the second research question. Although there were a number of “outlier” points in the data, the concentration of matched cases fell within the best-fit regression line. Thus, Figure 3 illustrated findings that the higher the Category 1: The Learner and Learning edTPA score, the higher the Category 1: The Learner and Learning CMSS score.

**Research Question 3.** To what extent do edTPA scores predict CMSS responses related to “InTASC Category 2: Content?”

To address the third research question, the researcher again matched the two data sources together using a participant-level unique identifier and cleaned the data, which resulted in n=106 complete records available for analysis. The researcher then grouped edTPA rubric scores and CMSS item values into InTASC Core Teaching Standards-aligned categories. Table 8 reports the results of a linear regression analysis of edTPA scores’ relationship with CMSS results related to InTASC Core Teaching “Category 2: Content” in relation to the research questions. Table 8 reports the descriptive statistics of estimate, standard error, t-value, p-value, and adjusted R2 for the linear regression analysis.

Table 8

*Linear Regression (#2): Use edTPA InTASC Category 2: Content Scores to Predict CMSS InTASC Category 2: Content Average Scores (n=106)*

<table>
<thead>
<tr>
<th>edTPA Category 2: Content</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t value</th>
<th>p value</th>
<th>Adj. R2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.010</td>
<td>0.010</td>
<td>0.928</td>
<td>0.356</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>

A linear regression was computed using edTPA Category 2 score as the single predictor variable and CMSS Category 2 average scores as the dependent variable. No statistically significant model was found. As presented in Table 6, the Category 2 regression model showed a
“trending” significance value of $p = .35$ resulting in a positive relationship of edTPA Category 2 predicting results on CMSS Category 2.

Figure 4 shows a scatter plot with associated best-fit line for the regression model tested.

Figure 4. edTPA Category 2: Content Scores and CMSS Category 2: Content Scores

Figure 4 displays the distribution of Category 2: Content edTPA scores on the x axis, and the y axis shows the values of Category 2: Content CMSS responses. The best-fit regression line (in red) demonstrated a slight upward trend, or positive relationship, between the two variables of the linear regression ($p = .35$). The relationship was not found to be significant. The plotted
data of Figure 4 illustrated a linear distribution of edTPA scores in relation to CMSS values. Regarding Category 2: Content, the best-fit regression line provides a visual representation of a slight positive correlation among the variables analyzed to answer the third research question. There was no distinct concentration of plotted data points found along the best-fit linear regression line. Thus, Figure 4 displayed findings that Category 2: Content edTPA scores were a “trending” positive predictor of Category 2: Content CMSS scores.

**Research Question 4.** To what extent do edTPA scores predict CMSS responses related to “InTASC Category 3: Instructional Practice?”

To address the fourth research question, the researcher matched the edTPA and CMSS data sources together using a participant-level unique identifier and cleaned the data, which resulted in n=106 complete records available for analysis. The researcher then grouped edTPA rubric scores and CMSS item values into InTASC Core Teaching Standards-aligned categories. Table 8 reports the results of a linear regression analysis of edTPA scores’ relationship with CMSS results related to InTASC Core Teaching “Category 3: Instructional Practice.” Table 9 reports the descriptive statistics of estimate, standard error, t-value, p-value, and adjusted R2 for the linear regression analysis.

Table 9

*Linear Regression (#3): Use edTPA InTASC Category 3: Instructional Practice Scores to Predict CMSS InTASC Category 3: Instructional Practice Average Scores (n=106)*

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t value</th>
<th>p value</th>
<th>Adj. R2</th>
</tr>
</thead>
<tbody>
<tr>
<td>edTPA Category 3:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructional Practice</td>
<td>0.010</td>
<td>0.007</td>
<td>1.632</td>
<td>0.106</td>
<td>0.016</td>
</tr>
</tbody>
</table>
A linear regression revealed a significance value of $p = .11$, thus demonstrating no strong evidence of a positive relationship for Category 3: Instructional edTPA scores predicting Category 3: Instructional Practice CMSS results.

Figure 5 shows a scatter plot with associated best-fit line for the regression model tested.

![Figure 5. edTPA Category 3: Instructional Practice Scores and CMSS Category 3: Instructional Practice Scores](image)

Figure 5 illustrates the distribution of Category 3: Instructional Practice edTPA scores on the x axis, and the y axis exhibits the values of Category 3: Instructional Practice CMSS responses. The best-fit regression line (in red) demonstrated a slight upward trend, or positive
relationship, between the two variables of the linear regression ($p = .11$). The relationship was not significant.

The plotted data of Figure 5 illustrated a generally scattered distribution of points, and a concentration of high CMSS values for a wide range of edTPA scores. Regarding Category 3: Instructional Practice, the best-fit regression line provided a visual representation of a slight positive correlation among the variables analyzed to answer the fourth research question. Thus, Figure 5 displayed findings that Category 3: Instructional Practice edTPA scores were not a significant predictor of Category 3: Instructional Practice CMSS scores.

**Research Question 5.** To what extent do edTPA scores predict CMSS responses related to “InTASC Category 4: Professional Responsibility?”

Addressing the fifth research question, the researcher again matched the two data sources together using a participant-level unique identifier and cleaned the data, which resulted in $n=106$ complete records available for analysis. The researcher then grouped edTPA rubric scores and CMSS item values into InTASC Core Teaching Standards-aligned categories. Table 10 reports the results of a linear regression analysis of edTPA scores’ relationship with CMSS results related to InTASC Core Teaching “Category 4: Professional Responsibility.” Table 10 reports the descriptive statistics of estimate, standard error, t-value, p-value, and adjusted R2 for the linear regression analysis.

Table 10

*Linear Regression (#4): Use edTPA InTASC Category 4: Professional Responsibility Scores to Predict CMSS InTASC Category 4: Professional Responsibility Average Scores ($n=106$)*

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t value</th>
<th>p value</th>
<th>Adj. R2</th>
</tr>
</thead>
<tbody>
<tr>
<td>edTPA Category 4:</td>
<td>0.057</td>
<td>0.031</td>
<td>1.866</td>
<td>0.065</td>
<td>0.023</td>
</tr>
<tr>
<td>Professional Responsibility</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A final linear regression showed $p = .06$, thus revealing no strong evidence of relationships pertinent to Research Question 5.

Figure 6 shows a scatter plot with associated best-fit line for the regression model tested.

Figure 6. edTPA Category 4: Professional Responsibility Scores and CMSS Category 4: Professional Responsibility Scores

Figure 6 shows the distribution of Category 4: Professional Responsibility edTPA scores on the x axis, and the y axis illustrates the values of Category 4: Professional Responsibility CMSS responses. The best-fit regression line (in red) demonstrated a slight upward trend, or positive relationship, between the two variables of the linear regression ($p = .06$). The relationship was not found to be significant.
The plotted data of Figure 6 shows a linear scattered distribution of points due to the concentration of values at the top of the scale for CMSS scores. Regarding Category 4: Professional Responsibility, the best-fit regression line provided a visual representation of a slight positive correlation among the variables analyzed to answer the fifth research question. Thus, Figure 6 displayed findings that Category 4: Professional Responsibility edTPA scores were not a significant predictor of Category 4: Professional Responsibility CMSS scores.

**Synthesis**

Two significant regression models were found. Overall edTPA score was found to be a significant predictor of overall CMSS average score: higher edTPA predicted higher CMSS. This lent support to Research Question 1. Of the four additional regression models computed, one model revealed a significant relationship where, again, higher edTPA scores organized into “InTASC Category 1: The Learner and Learning” predicted higher CMSS scores for InTASC Category 1. This finding provided good evidence for Research Question 2. Both of the statistically significant models estimated that edTPA accounts for around 3% of the variance in CMSS. This effect size was considered small.

The second of four secondary regressions showed a “trending” significance value of \( p = .35 \), leading to weak, mixed evidence of an effect between edTPA “InTASC Category 2: Content” and CMSS “InTASC Category 2: Content.” The Category 3: Instructional Practice regression model showed a significance value of \( p = .11 \), and the final regression model (Category 4: Professional Responsibility) showed \( p = .06 \). In sum, there was no strong evidence to show relationships pertinent to Research Questions 3, 4, or 5.
Summary

This chapter presented findings from an analysis of the relationship between 106 matched edTPA scores and CMSS results. Five research questions guided the study. The researcher presented linear regression results in tables and discussed for each of the five research questions. Chapter five presents an analysis of the results and provides recommendations for further study.
Chapter V: Discussion

Introduction

Understanding the effect of teaching quality on student learning by multiple measures is of paramount importance to educational stakeholders and the profession as a whole. Professional teaching standards guide educators’ current practice and inform continuous improvement efforts. The national InTASC teaching standards “outline the common principles and foundations of teaching practice that cut across all subject areas and grade levels and that are necessary to improve student achievement” (CCSSO, 2011, p. 3). Ensuring quality teaching in American schools that meets professional standards provides the foundation of educational accountability—particularly for teacher preparation programs throughout the United States (Cochran-Smith, 2001).

Multiple measures of teaching quality that are valid and reliable are essential in creating an environment of continuous improvement and reform of teacher education, as well as in-service support (Darling-Hammond, 2010). The Council for the Accreditation of Educator Preparation (CAEP) has established several criteria and sources of data on which Colleges of Education can demonstrate proficiency in meeting the standards (Council for the Accreditation of Educator Preparation, n.d.). Staff members in the College of Education at the selected university collect two assessments of teaching quality each year that are utilized for teacher preparation program improvement: the Teacher Performance Assessment (edTPA) and Common Metrics Supervisor Survey (CMSS).

The motivating issue for initiating the study was that the selected university utilized the edTPA and CMSS instruments independently. Specifically, the edTPA has been administered during student teaching, while the CMSS has been administered to supervisors at the end of the
new teacher’s first year of teaching, thus creating a gap in data collection. The purpose of the study was to examine the relationship between edTPA scores and perceived teaching quality at the end of the first year on the CMSS. The edTPA has been viewed as a valid and reliable performance assessment implemented nationally as a measure of teaching quality of teacher candidates prior to licensure and in-service teaching (SCALE, 2013b). The Common Metrics Supervisor Survey had “relatively good reliability” as a measure for evaluating a new teacher’s quality of instructional practice, understanding and work with diverse learners, quality of their classroom environment, and attending to professional responsibilities (Hezel Associates, 2013).

The researcher utilized the method of secondary data analysis of the edTPA and CMSS instruments. The edTPA and CMSS are both aligned to the national InTASC Model Core Teaching Standards (see Appendices B and C). Both CMSS and the edTPA are assessments collected by the small, public university in Minnesota. University staff collects teacher candidate edTPA data fall and spring semesters annually, and CMSS data are collected annually May-July. Data are utilized for non-research purposes at the university. The selected university stored candidate edTPA samples and edTPA results from past semesters. University staff also stored past CMSS data. The researcher acquired permission to analyze the university’s CMSS to compare those data with existing edTPA results for the university’s teacher preparation program completers. Identifying information was removed after matching the data from the CMSS and edTPA and verified using a TechID for analysis. Additionally, the researcher maintained validity and reliability of the edTPA and CMSS, as results from the actual administrations were used, and there were no revisions made to any aspects of the instruments.

To address the problem and purpose of the study, while drawing from the related literature, the researcher developed five research questions to guide the study. Due to the
importance of aligning outcomes to the professional teaching standards, the researcher organized
the research questions to align with the national InTASC Core Teaching Standards (CCSSO,
2011).

**Research questions.** Based upon the assertion that teaching quality is the most impactful
school-based factor on student learning (McCaffrey et al., 2003; Rivkin et al., 2000; Rowan et
al., 2002; Wright et al., 1997), and the presence of two existing data sets on Minnesota teacher
candidate teaching quality—edTPA and CMSS—the following research questions guided the
study:

1. To what extent do overall edTPA scores predict teaching quality at the end of the first
year as determined by supervisor responses on the Common Metrics Supervisor
Survey?
2. To what extent do edTPA scores predict CMSS responses related to “InTASC
Category 1: The Learner and Learning?”
3. To what extent do edTPA scores predict CMSS responses related to “InTASC
Category 2: Content?”
4. To what extent do edTPA scores predict CMSS responses related to “InTASC
Category 3: Instructional Practice?”
5. To what extent do edTPA scores predict CMSS responses related to “InTASC
Category 4: Professional Responsibility?”

**Research design.** In order to address the research questions, a criterion-referenced
predictive validity study was conducted. Linear regression analyses statistical tests were
employed to achieve the results discussed. The study utilized secondary data analysis
methodology to analyze the relationship between two data sources:

1. Education Teacher Performance Assessment (edTPA)
2. Common Metrics Supervisor Survey (CMSS)
The CMSS and edTPA are both aligned to the national InTASC teaching standards (see Appendices B and C). Both CMSS and the edTPA are assessments collected by the small, public university in Minnesota. Data are utilized for non-research purposes at the university, which places the survey and assessment in the existing data category. University staff has stored candidate edTPA samples and edTPA results from past semesters. University staff also has stored past CMSS data. CMSS data were collected with teacher preparation “program completer” names affixed, so CMSS data could be directly matched to the edTPA for comparative descriptive analysis. Analysis of the CMSS—a tool to assess supervisor perceptions of teaching quality after the first year of teaching—measured against the nationally valid and reliable edTPA performance assessment has revealed mixed results.

**Discussion and Conclusions**

For each of the five research questions discussed in depth below, the findings of the study revealed a slight-to-significant positive relationship between edTPA scores and CMSS scores. Results of the study will inform understanding and decision-making regarding teaching quality as assessed by the edTPA and CMSS for educational leaders in both higher education and K-12 education. More specifically, the findings of the study illustrated the ability of the edTPA to predict supervisor responses on the overall CMSS, as well as CMSS responses regarding learners and learning. In addition, although the relationship was determined to not be significant, there was a slight positive relationship between edTPA and CMSS regarding the InTASC categories of content, instructional practice, and professional responsibility.

**Research Question 1.** To what extent do overall edTPA scores predict teaching quality at the end of the first year as determined by supervisor responses on the Common Metrics Supervisor Survey?
Research Question 1 sought to ascertain the extent to which overall edTPA scores predicted overall CMSS results related to teaching quality. Analyzed as complete assessment tools of teaching quality, linear regression analysis determined that the edTPA was a significant predictor of CMSS average scores ($p = 0.04$). In other words, higher overall edTPA scores predicted higher overall CMSS responses. Those findings are important for the Institutions of Higher Education (IHE’s) that utilize both instruments to ascertain quality of their teacher candidates are of quality. Presently, the edTPA and CMSS are used independently as part of a system of multiple measures that provides data for program improvement.

As a result of the study’s significant findings regarding the overall assessments, educational leaders will have new knowledge about pre-service teaching quality. For example, IHE leaders, COE faculty and staff may be in a position to engage K-12 stakeholders in dialogue about the predictive ability of the edTPA in relation to supervisor perceptions of teaching quality. Based upon the findings of the study relating to research question one, K-12 educational leaders have more research-based information regarding the edTPA as a predictor of overall teaching quality after the first year of teaching. Thus, findings of the study might impact human resources decisions for the field. For example, K-12 educational leaders might ask potential hires about their achievement on the edTPA during the student teaching component of their preparation program. More pragmatically, K-12 educational leaders might be able to develop new teacher induction programs in order to improve teaching quality and retain quality teachers.

For leaders in higher education, research question one findings could impact the administrative scale-up of the CMSS to broadly understand the transition of teaching quality from pre-service to in-service work. Currently, CMSS administration is limited to a consortium of 14 IHE’s in the upper Midwest (Hezel Associates, 2013). With edTPA administration
occurring at a national level, educational leaders overseeing the CMSS have a research-based justification to expand CMSS administration as a measure of teaching quality. The Goldhaber et al. (2016) working paper described the edTPA’s lack of predictive ability for teaching quality. As a result of the study’s findings that overall edTPA scores significantly predict overall CMSS results, edTPA supporters might further support national CMSS administration as a response to the predictive ability critique of the edTPA.

**Research Question 1.** To what extent do edTPA scores predict CMSS responses related to “InTASC Category 1: Students and Student Learning?”

Research Question 2 sought to understand the extent to which edTPA scores predicted CMSS results related to students and student learning. Because both tools are aligned to the InTASC Core Teaching Standards, analyses were conducted to more specifically understand the relationship between the tools and codified national teaching standards. Research question two addressed InTASC Category 1: The Learner and Learning. Like the analysis of overall relationship between edTPA and CMSS, a linear regression analysis revealed that Category 1 edTPA scores significantly predicted Category 1 CMSS scores ($p = 0.036$). Thus, higher edTPA scores predicted higher CMSS responses regarding the learner and learning.

InTASC Core Standards Category 1: The Learner and Learning contained 10 out of 15 edTPA rubrics, as well as 23 of 45 survey questions on the CMSS. Combined, edTPA and CMSS assessment design addressed Category 1 more than InTASC Categories 2, 3, and 4. The study’s revelation of a significant positive relationship between the two instruments for InTASC Category 1 was essential for educational leaders’ understanding of both instruments.

**Research Questions 3, 4, and 5.** Research Questions 3, 4, and 5 sought to explain the extent to
which edTPA scores predicted CMSS results related to content knowledge, instructional strategies, and professionalism.

To what extent did edTPA scores predict CMSS responses related to “InTASC Category 2: Content?” To what extent did edTPA scores predict CMSS responses related to “InTASC Category 3: Instructional Practice?” To what extent did edTPA scores predict CMSS responses related to “InTASC Category 4: Professional Responsibility?”

Following the pattern of analyzing edTPA and CMSS based on the national teaching standards the researcher conducted three linear regression analyses to address the research questions relating to InTASC Category 2: Content, Category 3: Instructional Practice, and Category 4: Professional Responsibility. Linear regression analyses revealed there was not a statistically significant relationship of edTPA scores predicting CMSS results for InTASC Categories 2 (p = .35), Category 3 (p = .11), and Category 4 (p = .06) respectively.

Scatter plots revealed the distribution of supervisor responses relative to the value scale: not able to respond = 0, strongly disagree = 1, disagree = 2, agree = 3, strongly agree = 4. When plotted, CMSS responses for InTASC Categories 2, 3, and 4 were shown to be “higher” on the y axis of the scatter plots. Thus, it was reasonable to conclude that supervisors—more often than not for these three categories—rated their new teachers highly. The researcher concluded that because the four InTASC categories were much more likely rated “agree” and “strongly agree,” the regression analysis outcomes were affected. It was also reasonable to conclude, given the higher ratings of new teachers relative to InTASC Categories 2, 3, and 4, that supervisors who perceived success in those areas completed the CMSS more often than supervisors who may not have perceived success in those areas.
Limitations

Limitations are factors that occur after a study has commenced that may have affected the results (Roberts, 2010). Gaining approval to access the existing data files at the selected university became a primary limiting factor of the study. The university did not have a system in place to receive, review, and approve data requests for research studies. As a result, the approval process took several months for the specific College of Education faculty and leadership to develop a data request process.

Another limiting factor of the study was the number of cases the researcher would anticipate being included in the study. First, the presence of condition codes within edTPA data affected the statistical analysis used in the study, as the condition codes were valued at “zero” points. In addition, CMSS place holding values such as “999,” “998,” or “0” affected the legitimacy of those data for inclusion in the study. As a result, cases including either edTPA condition code and CMSS place holder data were excluded from the study, affecting the total number of cases included in the study.

The study’s final limiting factor was the reality that 100% of teacher candidates must submit an edTPA before completing their program, but for the CMSS to be administered, the teacher candidate must:

- Have been employed as a teacher.
- Have completed the Common Metrics Transition-to-Teaching Survey (TTS).
- Have provided a valid email address for their supervisor.
- Have their supervisor submit a CMSS.
Thus, four key factors affected the availability and legitimacy of CMSS data. This limiting factor reduced the total number of cases available for analysis in the study.

**Recommendations for the Field**

The study revealed that the overall edTPA was a significant predictor of overall CMSS average scores. Based on the linear regression model, every ten-point increase on the edTPA was associated with a 0.12-point increase on the CMSS average (on a 4-point scale). As a result, educational leaders are encouraged to consider the following recommendations for improving teaching quality in the education profession:

- Expand the administration of the CMSS nationally at institutions that also administer the edTPA. By expanding CMSS administration, College of Education educational leaders, faculty, and staff might engage in dialogue with K-12 education stakeholders regarding the edTPA’s predictive ability of teaching quality, thus improving the profession.

- Replicate the study at IHE’s that currently administer both the edTPA and CMSS in order to increase the number of cases (n=106 for the study). Also, replicating the study would also provide opportunities to better understand the predictive ability of the edTPA, as well as the understanding of the validity and reliability of the CMSS in relation to the InTASC Core Teaching Standards categories.

- Consider analyzing “teacher effectiveness” as a primary construct in better understanding the edTPA’s predictive ability related to overall CMSS responses. In other words, educational leaders are encouraged to pursue answering the important
question: “To what extent do new teachers who have completed the edTPA and CMSS impact student learning?”

- Educational leaders are encouraged to consider the degree to which they use multiple measures of teaching quality—edTPA and CMSS—in human resources decisions. For example, could educational leaders design research and data-based new teacher induction and professional development programs based upon edTPA and CMSS data?

**Recommendations for Future Research**

Results of the study and the related literature provide a framework for the following recommendations for future research:

- Future research may be conducted on the predictive ability of the edTPA relating to teacher effectiveness—this would involve student achievement as a variable.

- Related, future research may be conducted on the predictive ability of the edTPA of CMSS responses related to teacher effectiveness—this would require student achievement as a variable.

- An analysis of the impact of edTPA and CMSS as multiple measures of teaching quality on teacher preparation program improvement should be conducted.

- The current study may be extended to include participant demographics, such as race/ethnicity, gender, “low income flag,” grade point average (GPA), first-generation student, and other salient data collected by the university.
State education and accreditation agencies may consider researching legislative mandates on accountability in relation to the impact of qualitative and quantitative sources of data.

Summary

The study examined the relationship between the edTPA and CMSS as measures of teaching quality by a selected IHE. The profession has accepted and promoted the implementation of the edTPA as a valid and reliable assessment of teaching quality used for teacher licensure and program improvement (SCALE, 2013b). The CMSS is an emerging assessment tool and demands further study—especially its relationship to widely used tools evaluating teaching quality (such as the Danielson or Marzano framework). The purpose of the study was to examine the relationship between edTPA scores and perceived teaching quality after the first year of teaching as evaluated by the CMSS.

Teacher preparation program accountability drives improvements in teacher quality, teaching quality, and teacher effectiveness (Darling-Hammond, 2014; Lewis & Young, 2013). Deeply understanding the relationship between pedagogical performance during pre-service preparation and in-service teaching is critical to improving teaching and learning in 21st century American schools. Thus, the study’s findings have revealed a positive relationship of overall edTPA scores predicting overall CMSS results was statistically significant. Furthermore, the findings of the study proved to be professionally affirming in establishing that the edTPA and CMSS, as standards-based tools, are legitimate for assessing teaching quality in the pre-service and in-service phases of teacher preparation.

Conceptually, Shulman (1998) and, practically, Cochran-Smith (2001) advocated for educators to embrace their own future with regard to accountability. The study’s findings offer
educational leaders in both K-12 and higher education an opportunity to utilize significantly related teaching quality data to inform decisions affecting teaching and learning within their contexts. Educational leaders must seize the opportunity to embrace significant and meaningful data to ensure high quality teaching and learning for all. The story is far from over, however. As Shuman (1987) contended,

A knowledge base for teaching is not fixed and final. Although teaching is among the world’s oldest professions…it will, however, become abundantly clear that much, if not most, of the proposed knowledge base remains to be discovered, invented, and refined. (p. 12)
References


Grothendieck, G. (2014). SQLDF: Perform SQL selects on R data frames (R package version 0.4-10). Retrieved from https://CRAN.R-project.org/package=sqlfd


Appendix A: St. Cloud State IRB Approval Form

Institutional Review Board (IRB)
720 4th Avenue South MC 204K, St. Cloud, MN 56301-4498

Name: Joel J. Traver
Address: 5130 Highgrove Ln NW
          Rochester, MN 55901
          USA
Email: jjtraver@stcloudstate.edu

Project Title: An analysis of the relationship between Teacher candidate EdTPA scores and First-Year Teaching Quality.
Advisor: Dr. John Eller

The Institutional Review Board has reviewed your protocol to conduct research involving human subjects. Your project has been: APPROVED

Please note the following important information concerning IRB projects:
- The principal investigator assumes the responsibilities for the protection of participants in this project. Any adverse events must be reported to the IRB as soon as possible (ex. research related injuries, harmful outcomes, significant withdrawal of subject population, etc.).
- For expedited or full board review, the principal investigator must submit a Continuing Review/Final Report form in advance of the expiration date indicated on this letter to report conclusion of the research or request an extension.
- Exempt review only requires the submission of a Continuing Review/Final Report form in advance of the expiration date indicated in this letter if an extension of time is needed.
- Approved consent forms display the official IRB stamp which documents approval and expiration dates. If a renewal is requested and approved, new consent forms will be officially stamped and reflect the new approval and expiration dates.
- The principal investigator must seek approval for any changes to the study (ex. research design, consent process, survey/interview instruments, funding source, etc.). The IRB reserves the right to review the research at any time.

If we can be of further assistance, feel free to contact the IRB at 320-308-3290 or email irb@stcloudstate.edu and please reference the SCSU IRB number when corresponding.

IRB Institutional Official:

[Signature]
Dr. Latha Ramakrishnan
Interim Associate Provost for Research
Dean of Graduate Studies

OFFICE USE ONLY

<table>
<thead>
<tr>
<th>SCSU IRB# 1634 - 2045</th>
<th>Type: Exempt Review</th>
<th>Today's Date: 10/24/2016</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st Year Approval Date: 10/18/2016</td>
<td>2nd Year Approval Date:</td>
<td>3rd Year Approval Date:</td>
</tr>
<tr>
<td>1st Year Expiration Date:</td>
<td>2nd Year Expiration Date:</td>
<td>3rd Year Expiration Date:</td>
</tr>
</tbody>
</table>
Appendix B: edTPA Alignment to InTASC Core Teaching Standards (SCALE, 2013a)

<table>
<thead>
<tr>
<th>edTPA Tasks</th>
<th>edTPA Rubrics</th>
<th>InTASC Model Core Teaching Standards</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Category 1: The Learner and Learning (Standards 1-3)</td>
</tr>
<tr>
<td>Task 1: Planning</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1. Planning for Content Understandings</td>
<td>2, 3</td>
</tr>
<tr>
<td></td>
<td>2. Planning to Support Varied Student Needs</td>
<td>1, 2</td>
</tr>
<tr>
<td></td>
<td>3. Using Knowledge of Students to Inform Teaching and Learning</td>
<td>1, 2</td>
</tr>
<tr>
<td></td>
<td>4. Identifying and Supporting Language Demands</td>
<td>1, 2</td>
</tr>
<tr>
<td></td>
<td>5. Planning Assessments to Monitor and Support Student Learning</td>
<td>1</td>
</tr>
<tr>
<td>Task 2: Instruction</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6. Learning Environment</td>
<td>2, 3</td>
</tr>
<tr>
<td></td>
<td>7. Engaging Students in Learning</td>
<td>2, 3</td>
</tr>
<tr>
<td></td>
<td>8. Deepening Student Learning</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>9. Subject—Specific Pedagogy</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>10. Analyzing Teaching Effectiveness</td>
<td></td>
</tr>
<tr>
<td>Task 3: Assessment</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>11. Analysis of Student Learning</td>
<td></td>
</tr>
<tr>
<td></td>
<td>12. Providing Feedback to Guide Learning</td>
<td></td>
</tr>
<tr>
<td></td>
<td>13. Student Use of Feedback</td>
<td></td>
</tr>
<tr>
<td></td>
<td>14. Analyzing Students’ Language Use/Content Learning</td>
<td>1, 2</td>
</tr>
<tr>
<td></td>
<td>15. Using Assessment to Inform Instruction</td>
<td></td>
</tr>
</tbody>
</table>
### Appendix C: Common Metrics Supervisor Survey Alignment to InTASC Standards

<table>
<thead>
<tr>
<th>Associated InTASC Standard</th>
<th>Common Metrics Supervisor Survey</th>
</tr>
</thead>
<tbody>
<tr>
<td>InTASC 4</td>
<td>B1a. Effectively teach the subject matter in his/her licensure area.</td>
</tr>
<tr>
<td>InTASC 4</td>
<td>B1c. Design activities where students engage with subject matter from a variety of perspectives.</td>
</tr>
<tr>
<td>InTASC 1</td>
<td>B1d. Accounts for students’ prior knowledge or experiences in instructional planning.</td>
</tr>
<tr>
<td>InTASC 1</td>
<td>B1n. Uses digital and interactive technologies to achieve specific learning goals.</td>
</tr>
<tr>
<td>InTASC 1</td>
<td>B1p. Helps students develop critical thinking processes.</td>
</tr>
<tr>
<td>InTASC 1</td>
<td>B1q. Helps students develop skills to solve complex problems.</td>
</tr>
<tr>
<td>InTASC 1</td>
<td>B1s. Knows where and how to access resources to build global awareness and understanding.</td>
</tr>
<tr>
<td>InTASC 1</td>
<td>B1t. Helps students analyze multiple sources of evidence to draw sound conclusions.</td>
</tr>
<tr>
<td>InTASC 2</td>
<td>B2a. Effectively teaches students from culturally and ethnically diverse backgrounds and communities.</td>
</tr>
<tr>
<td>InTASC 2</td>
<td>B2b. Plans differentiated instruction for a variety of learning needs.</td>
</tr>
<tr>
<td>InTASC 2</td>
<td>B2c. Uses developmentally appropriate practices to support student learning.</td>
</tr>
<tr>
<td>InTASC 2</td>
<td>B2d. Addresses the needs of students from various socioeconomic backgrounds.</td>
</tr>
<tr>
<td>InTASC 2</td>
<td>B2e. Designs instruction for students with IEPs and/or 504 plans.</td>
</tr>
<tr>
<td>InTASC 2</td>
<td>B2f. Designs instruction for students with mental health needs.</td>
</tr>
<tr>
<td>InTASC 2</td>
<td>B2g. Designs instruction for gifted and talented students.</td>
</tr>
<tr>
<td>InTASC 2</td>
<td>B2h. Designs instruction for English language learners.</td>
</tr>
<tr>
<td>InTASC 2</td>
<td>B2i. Accesses resources, programs, and other school personnel to foster student learning.</td>
</tr>
<tr>
<td>InTASC 8</td>
<td>B1b. Selects instructional strategies to align with learning goals and standards.</td>
</tr>
<tr>
<td>InTASC 8</td>
<td>B1o. Engages students in using a range of technology tools to access, interpret, evaluate, and apply information.</td>
</tr>
<tr>
<td>InTASC 8</td>
<td>B1r. Makes interdisciplinary connections among core subjects.</td>
</tr>
<tr>
<td>InTASC 3</td>
<td>B3c. Designs instruction and learning tasks that connect core content to real-life experiences for students.</td>
</tr>
<tr>
<td>InTASC 3</td>
<td>B3d. Helps students work cooperatively to achieve learning goals.</td>
</tr>
<tr>
<td>InTASC 3</td>
<td>B3e. Develops and maintains a classroom environment that promotes student engagement.</td>
</tr>
<tr>
<td>InTASC 3</td>
<td>B3g. Creates a learning environment in which differences such as race, culture, gender, sexual orientation, and language are respected.</td>
</tr>
<tr>
<td>InTASC 3</td>
<td>B3h. Uses classroom management techniques that foster self-control and self-discipline among students.</td>
</tr>
<tr>
<td>----------</td>
<td>-------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>InTASC 3</td>
<td>B3i. Effectively organizes the physical environment of the classroom for instruction.</td>
</tr>
<tr>
<td>InTASC 3</td>
<td>B3a. Clearly communicates expectations for appropriate student behavior.</td>
</tr>
<tr>
<td>InTASC 5</td>
<td>B3b. Uses effective communication skills and strategies to convey ideas and information to students.</td>
</tr>
<tr>
<td>InTASC 3</td>
<td>B3f. Responds appropriately to student behavior.</td>
</tr>
<tr>
<td>InTASC 7</td>
<td>B1e. Designs long-range instructional plans that meet curricular goals.</td>
</tr>
<tr>
<td>InTASC 7</td>
<td>B1f. Regularly adjusts instructional plans to meet student needs.</td>
</tr>
<tr>
<td>InTASC 7</td>
<td>B1g. Plans lessons with clear learning objectives/goals in mind.</td>
</tr>
<tr>
<td>InTASC 6</td>
<td>B1h. Designs and modifies assessments to match learning objectives.</td>
</tr>
<tr>
<td>InTASC 6</td>
<td>B1i. Provides students with meaningful feedback to guide next steps in learning.</td>
</tr>
<tr>
<td>InTASC 6</td>
<td>B1j. Engages students in self-assessment strategies.</td>
</tr>
<tr>
<td>InTASC 6</td>
<td>B1k. Uses formative and summative assessments to support student learning.</td>
</tr>
<tr>
<td>InTASC 6</td>
<td>B1l. Identifies issues of reliability and validity in assessment.</td>
</tr>
<tr>
<td>InTASC 6</td>
<td>B1m. Uses multiple and appropriate types of assessment data to identify student learning needs.</td>
</tr>
<tr>
<td>InTASC 6</td>
<td>B2j. Develops fair and unbiased assessments for all learners.</td>
</tr>
<tr>
<td>InTASC 9</td>
<td>B4a. Seeks out learning opportunities that align with my professional development goals.</td>
</tr>
<tr>
<td>InTASC 9</td>
<td>B4b. Accesses the professional literature to expand my knowledge about teaching and learning.</td>
</tr>
<tr>
<td>InTASC 10</td>
<td>B4c. Actively engages with parent/guardian/advocate about issues affecting student learning.</td>
</tr>
<tr>
<td>InTASC 10</td>
<td>B4d. Collaborates with teaching colleagues to improve student performance.</td>
</tr>
<tr>
<td>InTASC 10</td>
<td>B4e. Uses colleague feedback to support my development as a teacher.</td>
</tr>
<tr>
<td>InTASC 10</td>
<td>B4f. Upholds my legal responsibilities as a professional educator and student advocate.</td>
</tr>
</tbody>
</table>
Appendix D: edTPA Reliability (SCALE, 2013b, p. 24)

<table>
<thead>
<tr>
<th>Task</th>
<th>Rubric</th>
<th>Agreement Rate</th>
<th>Kappa - N</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Task 1: Planning</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rubric 01</td>
<td></td>
<td>0.929</td>
<td>0.852</td>
</tr>
<tr>
<td>Rubric 02</td>
<td></td>
<td>0.832</td>
<td>0.65</td>
</tr>
<tr>
<td>Rubric 03</td>
<td></td>
<td>0.934</td>
<td>0.863</td>
</tr>
<tr>
<td>Rubric 04</td>
<td></td>
<td>0.931</td>
<td>0.856</td>
</tr>
<tr>
<td>Rubric 05</td>
<td></td>
<td>0.872</td>
<td>0.733</td>
</tr>
<tr>
<td><strong>Task 2: Instruction</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rubric 06</td>
<td></td>
<td>0.962</td>
<td>0.921</td>
</tr>
<tr>
<td>Rubric 07</td>
<td></td>
<td>0.943</td>
<td>0.881</td>
</tr>
<tr>
<td>Rubric 08</td>
<td></td>
<td>0.932</td>
<td>0.858</td>
</tr>
<tr>
<td>Rubric 09</td>
<td></td>
<td>0.868</td>
<td>0.725</td>
</tr>
<tr>
<td>Rubric 10</td>
<td></td>
<td>0.946</td>
<td>0.887</td>
</tr>
<tr>
<td><strong>Task 3: Assessment</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rubric 11</td>
<td></td>
<td>0.911</td>
<td>0.815</td>
</tr>
<tr>
<td>Rubric 12</td>
<td></td>
<td>0.934</td>
<td>0.863</td>
</tr>
<tr>
<td>Rubric 13</td>
<td></td>
<td>0.925</td>
<td>0.844</td>
</tr>
<tr>
<td>Rubric 14</td>
<td></td>
<td>0.941</td>
<td>0.877</td>
</tr>
<tr>
<td>Rubric 15</td>
<td></td>
<td>0.906</td>
<td>0.804</td>
</tr>
<tr>
<td><strong>Overall</strong></td>
<td>Average</td>
<td>0.917</td>
<td>0.829</td>
</tr>
</tbody>
</table>

**edTPA Rater Agreement Rates.** This table presents rater agreement rates (exact plus adjacent agreement rates) and \( K_n \) (Kappa-N) agreement rates for the 2013 edTPA field test.