View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by D-Scholarship@Pitt

What Matters? An Analysis of Teacher Training, Student Achievement, and Teacher Attrition in North Carolina.

by

Kimberlee Eberle-Sudré

Bachelor of Arts, Wittenberg University, 2003

Masters of Science, University of Kansas, 2005

Submitted to the Graduate Faculty of the

School of Education in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2020

UNIVERSITY OF PITTSBURGH

SCHOOL OF EDUCATION

This dissertation was presented

by

Kimberlee Eberle-Sudré

It was defended on

May 26, 2020

and approved by

Dr. M. Najeeb Shafiq, Professor, Administrative and Policy Studies

Dr. Lindsay Page, Associate Professor, Psychology in Education

Dr. Patricia Crawford, Associate Professor, Instruction and Learning

Dr. Karen Goldman, Adjunct Instructor, Chatham University

Dissertation Advisor: Dr. Sean Kelly, Professor, Administrative and Policy Studies

Copyright © by Kimberlee Eberle-Sudré

2020

What Matters? An Analysis of Teacher Training, Student Achievement, and Teacher Attrition in North Carolina.

Kimberlee Eberle-Sudré, PhD

University of Pittsburgh, 2020

Using a quantitative approach that draws on qualitative methods, this study aimed to identify what, if any, elements within an elementary teacher preparation program have a relationship between elementary student achievement gains in reading and math as well as a teacher's attrition. This study identified elements of teacher training programs through interviews, program document reviews, and surveys of training programs in North Carolina. I used data from the North Carolina Education Resource Data Center (NCERDC) for the years 2014-2018 to test relationships between program elements, student achievement, and teacher attrition using a covariate adjusted multilevel linear model and a school fixed effects model.

Results showed that attendance at certain training programs themselves can increase student achievement in math and reading, as well as a teacher's attrition. Furthermore, there were signs that specific courses in the training program (such as subject matter) can also impact a student's learning and teacher attrition. Specifically, there were two elements of training techniques themselves that showed an impact on student achievement and teacher attrition: recording a candidate's teaching practice in student teaching experiences and lessons on instructional technology. Such findings suggest elements within training likely matter as much as the overall training requirements and more research is needed within programs to test the effects of elements themselves, as well as the quality of the element on student and teacher outcome.

Table of Contents

List of Tables

List of Figures

List of Equations

Preface

Ten years ago, I was sitting in the back of a room, a naïve, yet hungry, graduate intern at the US Department of Education. A policymaker randomly asked me if I had time to do some analysis on teacher training policies for federal regulations they were considering. I, knowing nothing of the topic besides my own individual teacher training experience, said “Sure!”. Little did I know that would set the stage for my research as well as the rest of my career. I am greatly appreciative for that moment, but there are so many people who got me to that room and who have since been guides along the way.

Dr. Kelly. There are no words to truly express my gratitude. Had you put myself as a data point into a probability equation of “chances of getting to the finish line”, you likely wouldn’t have taken me on as a student. I am so thankful you knew to look beyond probabilities as some things, such as a student’s grit, cannot be easily measured. Thank you for your patience, your guidance, and for taking a risk on me. I am forever grateful for your unwavering support.

Dr. Shafiq. At one point in my doctoral career, you pulled me aside and told me that during the dissertation process there will be times that may feel dark and isolating. You reminded me to look past those moments and to remember there is light at the end. Like all those before me, with slow and steady dedication, I would make it to the end. I carried your words with me daily through this process. That I got here is, in part, because of your words. Thank you for believing in me, even when I wasn’t fully able to believe in myself.

Dr. Page. You always asked me the hard questions. Those questions I hoped no one would ask, but I knew you would ask. You pushed me to dig deeper, to question more, and to never shy away from the unknown. As a result, my research is that much stronger. Thank you.

Dr. Crawford, thank you so much for your thoughtful words and willingness to talk at length about the intricacies of teacher training models and techniques. I always left our conversations with more insight and a deeper appreciation of the complexities of teacher training. Thank you.

Karen, I am so grateful I got to work with you. You are the one who trained me to teach. You have been my cheerleader every step of the way. Thank you.

Dr. McClure, I was at a crossroads between a field study experience in Brazil and an internship in DC. Without hesitation, you said, “Go to DC. Any other questions?” You not only helped me make a decision that would forever change my path, but also gave me advice to ensure I could ‘survive DC’. Thank you.

I am forever indebted to those I interviewed and those who participated in the survey for this study. It is because of your work that this study matters. I hope some part of this study is helpful in your work. I’d also like to thank Kara Bonneau at the NCERDC for helping me understand my data and for answering even the simplest of questions.

My first ‘family’ at the University of Pittsburgh, those at the Latin American Center. You not only provided me with funding for much of my time at the University but also with a community. I deeply value my time with you and the lessons learned while a part of the Center. Additionally, I am greatly appreciative of my support from the School of Education, without which I would not have been able to complete this study.

To those at the Association of American Universities, you quite literally got me over the finish line. I am so thankful for the experiences you have given me that allowed me to grow personally and professionally. And, a very special thank you to Toby, for your encouragement, guidance, understanding, and the most precious of all: the gift of time.

Jen, Shang, and Heejin. My Pitt cohort. Each of you went before me and taught me small lessons along the way that ensured I could get to the end. Most importantly, you showed me kindness and friendship during the hardest of times. Thank you.

I pursued most of this work away from the brick and mortar walls of my university. As such, I built an unofficial cohort of colleagues who provided unwavering empathy. Many thanks to Tara, Emily, Hiro, and Konrad. And Meredith, my former work wife, you've heard about this study since before I even had a real question. Thank you for listening, for asking questions, and for never rolling your eyes even though we had the same conversation every day.

A minha família, Gustavo e Stella. Vocês são o meu coração, a minha alma, e a minha vida. Eu amo vocês com o meu tudo. Sem vocês eu não teria conseguido alcançar mais essa meta. Obrigada por tudo.

Gustavo, there is no other person I can think of who would willingly read a multilevel modeling textbook just to proofread their partner's dissertation and confirm "yes, that paragraph makes sense". But as I've always known, there is only one you. Te amo amoreco, te amo.

Stella "Loops". If anyone had told me that having a child would light a fire inside me to finish this study, I would never have believed them. Luckily, you came into our lives and could be a part of this. Your laughter and constant energy have helped me continue on this path, for which I am eternally grateful. And yes, Stella, now I would love to build a snowman with you.

For Loops~

If you dream it, believe in it, and work hard for it, anything is possible.

1.0 Introduction

There is wide agreement among researchers that teachers are the single-most important in-school factor to affect a student's achievement (Aaronson, Barrow, & Sander, 2007; King, 2010; Nye, Konstantopoulos, & Hedges, 2004; Rivkin, Hanushek, & Kain, 2005; Rockoff, 2004). Indeed, the benefits of high teacher quality have been tied to significant increases in student learning: an estimated 1.5 years in student achievement gains (which have subsequently led to a 10-20 percent increase in financial earnings in a student's lifetime) (Chetty, Friedman, & Rockoff, 2011; Hanushek & Rivkin, 2010).

Such an impact is noteworthy, given the size and scope of the teaching profession. The most recent data available shows that teaching is the largest profession in the public sector (US Department of Labor, 2017). In 2016, there were 3.8 million teachers in public schools throughout the US (approximately 1.9 million in elementary and 1.9 million in secondary schools), who were responsible for teaching the 47.2 million k-12 students in public schools (US Department of Education, 2016a)[1]. Furthermore, researchers estimate that the US k-12 student population will grow by approximately 1 million students in the next decade (with most gains occurring in the public sector), resulting in an increased need of more than 200,000 teachers in public schools (Hussar & Bailey, 2018).

When looking at the vast size of the teaching profession, along with the increased demand for teachers over the next ten years, it is important to also consider the supply and 'production' of

[1] Recent data is not available for those in private k-12 schools, but based on numbers from 2012, there are estimated to be less than 600,000 teachers total (US Department of Education, 2015b).

teachers. As of 2016, approximately 80 percent of all public-school teachers were trained in a traditional teacher training program which was housed at an institution of higher education (and 90 percent of all elementary teachers were trained through this route) (US Department of Education, 2016a). Although degrees in education have declined by 14 percent since 2010, these programs are still the biggest contributor of teachers in the public-school system (AACTE, 2018). Furthermore, national survey data of recently trained teachers suggest that eight percent of teachers leave the profession completely after five years (US Department of Education, 2015a). Such retention patterns in teaching amplify the need for even more teachers who are trained and prepared to enter the classroom. Given the increasing demand for teachers, the vast size of the profession, and the positive effect teachers can have on student outcomes, it is not surprising that policymakers (at federal and state levels) have sought to ensure the supply of well-trained teachers meets the demand when they enter their classrooms for the first time.[2]

While researchers, policymakers, and training programs themselves have pushed for effective training, what makes the training effective and in turn makes a teacher effective, is a hotly debated topic. For instance, researchers have debated approaches and models to evaluate a program's effects, the federal government has proposed regulations to change current practices and data reporting on training, states have implemented vast reforms throughout programs that they oversee, and programs themselves are examining their program offerings and training techniques (Lewis & Young, 2013). Most of these efforts that tie teacher training and its effects

[2] For purposes of this paper, teacher quality and teacher effectiveness are defined as the ability of teachers to meaningfully influence student achievement gain scores as measured by standardized tests. While there are notable limitations to this narrow definition, it is employed because the method of measuring student outcomes in this study will be based on achievement scores from North Carolina state exams.

on student achievement however, have questioned whether training programs overall make teachers more effective, and if so, which programs have the highest outcomes. Few of these efforts have explored specific elements *within* individual training programs and their relationship to student outcomes. This is primarily due to the limited availability of administrative datasets that link teacher training to teachers and subsequently to student achievement scores. Furthermore, historically, program elements have not varied substantially enough to measure differences in student outcomes that result from a teacher's training. However, vast data improvements have occurred over the last decade and program types and program training techniques have diversified (Lewis & Young, 2013).

One state in particular, North Carolina, is a ripe research ground for such an examination. With more than 1.5 million students enrolled in the public k-12 school system and almost 100,000 teachers, North Carolina has the ninth largest student enrollment in the United States as of 2016 (US Department of Education, 2015b). In addition to its large student enrollment, North Carolina also has a vast data system that connects student records (including student achievement records and demographic data of those in public schools), to teacher records (including teacher credentials, education, and demographics), as well as employment data that goes back to the early 2000's. This robust data system was originally built in 1992 to connect state-wide agency data to outcomes on educational, employment, and training programs (North Carolina Department of Public Instruction [NCDPI], 2016). Since then, the effort has continually expanded to include pre-school student records, IRS data on student and employee wages, and data on university student outcomes. The state now has one of the largest longitudinal data systems of students and employees in the nation, and due to the wide infrastructure in place, it is also available to researchers not associated with a North Carolina agency (NCDPI, 2016). Through the use of this data system, North Carolina has also become a leader in using data to drive reforms, especially those in teacher training. For instance, North Carolina was one of the first states to longitudinally track student outcomes and teacher training. Due to insights gleaned from this data, training programs within the state have made changes to their training requirements, techniques, and overall program structures (Bryan, Hougen, & Nelson, 2018).

To add to the current literature on the effects of teacher training, this study will use the rich data provided by the North Carolina Educational Research Data Center (NCERDC) to identify what, if any, elements within teacher training programs contribute to elementary student achievement and teacher attrition. Because of the dynamic environment surrounding teacher education however, this chapter will first define the landscape of teacher education at the research, federal, state, and local (program) levels. The first section of this chapter will outline the academic research around teacher training, noting trends in findings, the debates within those studies, and promising areas yet to be explored. This chapter will then turn to the political debates and governmental regulations surrounding teacher training at higher education institutions. In recent years, the federal and state governments have proposed (and in many cases passed) legislation aimed to regulate teacher education and have impacted training practices at many programs (Lewis & Young, 2013). This section aims to highlight those governmental efforts and the changes experienced within programs. Noting this changing environment, this chapter will then examine recent reforms in teacher education within the state of North Carolina and highlight research efforts, state specific regulations, as well local training efforts aimed to improve teacher training within the state. This chapter ends with a brief outline of my study and notes how the study fits into the research, federal, and state level contexts of teacher training.

1.1 Research Context

Research on the impacts of teacher preparation and effective training methods on a teacher's quality and student achievement have grown over the last 20 years, increasing as the quality of the data improved, statewide data systems were developed, and newer data collection and linkage efforts have been employed. However, there is still limited agreement among researchers about what elements of teacher training are the most impactful. Perhaps more concerning is that there is a wide debate about whether training even impacts a teacher's effectiveness (Henry, Purtell, et al., 2014). One of the first studies on teacher training effects was in 2009. Authors Boyd, Grossman, Lankford, Loeb, & Wyckoff (2009) examined longitudinal data on New York City teachers, their training, and student achievement outcome data and found conclusive variation between the effectiveness of teachers graduating from different programs. Authors noted that "a change of one standard deviation in the effectiveness of a training program corresponds more than a third of a standard deviation in the teacher effect for new teachers" (p.429). Findings from this study (and the availability of similar datasets throughout states), set off a large debate amongst researchers questioning methods and metrics used to determine what, if anything, actually matters in teacher training.

Researchers Goldhaber & Liddle (2011) built off of Boyd et. al's (2009) model and used a sample of teachers and students in Washington state, and similarly found modest, yet significant, results. Their study found that the "standard deviation of the program effect is 18 -20 percent of the standard deviation of the teacher effects (for math and reading, respectively)" (Goldhaber & Liddle, 2011, p. 17). However, a study by Koedel, Parsons, Podgursky, & Ehlert (2012) challenged these findings and argued that an inaccurate research design led to an overestimate of findings. Authors argued that real variation existed between different training programs within a single institution and not between institutions. That is, one institution may have several different training programs such as elementary education, secondary science, or secondary math, and these programs may vary in quality (thus one institution may graduate teachers from different programs that vary in quality). They also argued that while there are significant differences between some programs, these differences are very small (with an effect of approximately one percent of a standard deviation) and as a result cannot be used as a measure of program quality. Goldhaber, Liddle, & Theobald (2013) responded to these findings with a follow-up study that used a similar dataset from Washington state but added Koedel et al's., (2012) model corrections. Using the revised model, authors still found significant variations between programs (though lower than previously), as the "standard deviation of program estimates was between 5-12 percent of a standard deviation of the teacher effects" (Goldhaber et al., 2012, p. 36). Such findings led researchers to argue that while there may be some effect on student achievement due to training, the effect size is often because of method specifications.

Researchers have continued this line of inquiry throughout states with similarly small, yet significant findings in the differences between programs. For instance, using student data from North Carolina, Henry, Purtell et al., (2014) applied two different statistical models (a covariate adjusted multilevel model and a school fixed effects model) to examine the differences in teacher effectiveness based on the training route that the teacher used to enter the profession (such as traditional in-state university, alternative training, out-of-state traditional university, etc.). They found small differences among some training programs. For instance, teachers trained at programs housed within public universities in North Carolina were more effective than other entry routes in 14 tested subjects/grades (however, findings were small with the largest effect being 0.049 standard deviations in high school math).

Authors von Hippel, Bellows, Osborne, Lincove, & Mills (2016) noted that when testing teacher training effects on student achievement, research decisions can often account for the variation in findings. For instance, decisions on which covariates to include, whether to add school fixed effects, as well as testing a random effects model have all shown a broad range of findings. In their study, von Hippel et al., (2016) tested these varying approaches on a state-wide database for the state of Texas. They found that differences existed between programs, but they were small (0.04 standard deviations for math scores and 0.03 standard deviation in reading). However, they also noted that these estimates were noisy, and they questioned the reliability of their findings as each statistical decision (such as applying random or fixed effects) saw wide variations in results (as some programs had positive effects in one model and negative effects in another). In a follow-up study, von Hippel & Bellows (2018) examined six different studies on teacher training and student achievement from six different states and tested differing research methodologies to see if they could replicate the original author's findings when statistical corrections and modifications were made. This study showed that once corrections were made, there were marginal differences in a teacher's quality depending on the program from which she was trained (with a 0.01-0.03 increase in standard deviations in student test scores), even though some of the studies they tested found large differences amongst programs. They noted that some statistical methodologies make program effects appear larger than they are (such as clustered standard errors at the teacher level, using an experienced teacher comparison group, and applying random effects at the teacher level). However, they did confirm that although modeling can influence the size of effect, there were some programs that had large, statistical differences regardless of the model used.

While the debate continues as to whether programs themselves have an overall effect on a teacher's effectiveness, researchers have begun to dig within programs to ask if there are components within individual programs that may be responsible for teacher and student outcomes. Although results have been mixed, there are promising findings that subject specific courses (such as mathematics) and courses on how to teach a specific subject (such as how to teach mathematics) have a positive association with student achievement and teacher effectiveness. For instance, researcher Kukla-Acevedo (2009) examined training program effects on fifth-graders in Kentucky and found that a teacher's number of course hours in both math content as well as in teaching math, had a positive relationship to a student's math achievement. In a similar study, authors Harris & Sass (2011) used longitudinal data on students and teachers in Florida to compare professional development modules as well as traditional teacher training programs. They found that content-specific professional development courses were positively related to a teacher's effect. Furthermore, they found that content specific courses within traditional training programs were also related to a teacher's effectiveness (though were only present at the middle and high school levels). Boyd et al., (2009) in their study of training programs of teachers in New York City, also noted that teachers who were trained at programs with more content courses had higher student math scores. One unique finding from this study was that candidates who were placed in clinical experiences that were similar to their first job placement also had higher student achievement scores.

These findings, however, differed slightly from those of Henry et al., (2013). In this study, the authors examined whether teacher candidate characteristics at one North Carolina training program (such as GPA, clinical assessments, portfolio assessments etc.) saw any difference in student achievement. They found that while overall performance and number of content courses had a positive influence on student test scores, measures of a teacher's clinical experience, individual grades in core courses, and other measures of performance did not influence a teacher's effectiveness. Preston (2017) furthered Henry et al.'s (2013) work and looked at the effects of teacher training programs (at all public universities in North Carolina) and the elements within those programs to determine an effect of training on a student's achievement. This study found that too many content-specific courses had a negative relationship to student achievement in middle grades. Furthermore, there was a negative relationship between early teaching experiences and student math achievement, suggesting that candidates may not be learning content pedagogical methods in these courses and experiences.

Research suggests that the effects of teacher training programs on teacher quality, and ultimately student achievement, is limited and very small (when significant). Furthermore, statistical and methodological decisions are often responsible for these differences. However, all studies have found that there are programs with significant effects and have noted these differences are worth exploring (von Hippel & Bellows, 2018). As such, many have begun to dig deeper into programs to determine what may contribute to a program's effectiveness. Researchers have questioned the clinical experience, the number of content course hours required by the program, as well as the number of courses on teaching content. Researchers have found some promising results such as a positive relationship between content specific courses and content pedagogy and student outcomes. However, as the data is limited on program requirements and this area of research is new, fewer studies have been able to link program elements to teachers and then to student achievement (Preston, 2017).

1.2 Federal Context

The debate on the effectiveness of teacher training is not limited to researchers, however. Indeed, policymakers have also furiously debated how to improve teacher training, arguing the stakes are too high to not improve teacher training. This came to a head in 1998, when the United States Congress sparked a heated debate on education accountability and transparency through the amendments of the Higher Education Act (HEA). Under Title II of the HEA of 1998, Congress required that states report on each institution of higher education that received federal funding for student financial aid and was approved to prepare students to teach (and issue state teaching licenses) (Higher Education Opportunity Act, 2008). States were required to publicly release report cards with evidence of program success such as: percentage of students that complete coursework, percentage of students that pass licensure examinations and their scores, and students who enroll and pass through all steps of the program. States were also required to publicly list programs rated ‘low-performing’ as determined by a state defined criterion (Higher Education Opportunity Act, 2008). Failure to report on any of these issues would result in a $25,000 fine, and programs that were rated as low performing risked losing federal student financial aid (Higher Education Opportunity Act, 2008).

Congress and the US Department of Education argued that such transparency measures were necessary to a) ensure that students were informed of program outcomes b) apply more accountability measures to teacher preparation programs and c) improve teacher training for the betterment of US education (US Department of Education, 2003). Starting in 1998, the reporting regulations for teacher preparation programs were implemented and the first state report cards were published (US Department of Education, 2003). Soon after, the US Department of Education began to question the validity of the information reported and whether the report cards actually succeeded in pressuring ineffective programs to improve (Crowe, 2011). For instance, the US Department of Education (2011) noted that in 2010, only 37 teacher preparation programs (two percent of all eligible programs) were designated as ‘low-performing’. Each state was allowed to define criteria for ineffective and effective programs, as well as licensure exam pass rates. This led states to allow teacher candidates to retake the required licensure exams until the candidate passed (without reporting the number of times a candidate took the exam) resulting in overwhelmingly high passage rates. Furthermore, many training programs required candidates to pass the exam to graduate, thus both graduation and exam passage rates were above 95 percent (Aldeman, Carey, Dillon, Miller, & Silva, 2011). Many argued that licensure completion and program completion data were not accurate measures of a teacher’s classroom preparation, but more the ability to pass the licensure exam.

During this same period, as policymakers were examining the effectiveness of teacher training, researchers also questioned new teacher’s feelings of preparedness. Specifically, Levine’s (2006) survey of new teacher’s preparedness found only 38 percent of first year teachers felt adequately prepared to enter the classroom. Levine (2006) attributed this lack of preparation to limited preparation in classroom teaching experience as only 50 percent of programs provided candidates with supervised clinical training. Additionally, districts were becoming concerned because reportedly one-third of new teachers quit the profession after three years, leaving districts to continually recruit teachers (and estimated costs of this continual recruitment was estimated to be $7 billion in 2007 to recruit, retrain, and retain teachers) (Barnes, Crowe, & Schaefer, 2007). Furthermore, districts noted that the profession was oversaturated with elementary teachers, yet districts were unable to fill positions in the areas of: English as a second language, math, science, and special education (Ingersoll & Perda, 2009).

Citing these concerns, the US Department of Education and the US Congress[3] proposed regulations and legislation to require states to publicly report teacher value-added scores linked to teacher training programs (Sawchuck, 2013a). Arguments in support of this increased information claimed such data would provide consumers (and students) with more information and could ultimately lead to the improvement of lower ranked programs because aspiring teachers would not want to attend a training program where graduates were not effective (Crowe, 2011). These regulations also proposed penalizing institutions that housed programs that didn't comply through fines and in extreme cases elimination of student federal aid eligibility for students within those programs (US Department of Education, 2016d). However, federal efforts to reform teacher education were not limited to regulatory and legislative efforts. In 2010, through the federal grant competition, Race to the Top, the US Department of Education also sought to incentivize state reform efforts by providing financial awards for states that used program and student achievement data to inform teacher training program reforms (Henry, Kershaw, Zulli & Smith, 2012; US Department of Education, 2010).

Training programs, however, argued that these requirements had unfair consequences because they threatened the loss of student financial aid and increased reliance on licensure exams, which programs argued were not the best measure of program quality (Tatto, et al., 2016). Concerns arose that these reporting requirements could cause financial consequences that would lead to the closure of programs due to decreased enrollment from a) public shaming if labeled

[3] Both the US House of Representatives and the US Senate proposed bills in 2013 (H.R. 2172/S.1062 Educator Preparation Reform and H.R. 2173/S. 1063 Effective Teaching and Leading Acts) related to the proposed regulations from the US Department of Education (Sawchuck, 2013).

ineffective/low performing or b) reduced student aid if a program was noncompliant or ineffective (Tatto, et al., 2016). Ultimately, due to significant concerns from teachers, training programs, and institutions of higher education, these regulations were not adopted and congressional legislation was not enacted (Tatto, et al., 2016). Although federal efforts were not implemented, the Governmental Accountability Office (GAO) issued a report of federal and state oversight of teacher training programs and noted that the current systems in place to evaluate teacher training programs were insufficient. While several states were implementing significant reforms, those states received limited support from the federal government as the US Department of Education did not have mechanisms in place to share data and evidence on effective reform efforts (United States Government Accountability Office, 2015).

During this same period that the US federal government was debating regulations on teacher training programs, lawmakers were also reauthorizing the 1965 Elementary and Secondary Education Act (ESEA). This reauthorization, the Every Student Succeeds Act (ESSA), was passed in 2015 and sought to reduce federal oversight of state accountability systems that had previously been present in prior ESEA legislation as well as in federal programs such as Race to the Top (Close, Amrein-Beardsley, & Collins, 2018). Through this legislation, states were permitted to adopt new models of school accountability, new teacher evaluation systems that were not tied to former value-added models, and were given the opportunity to use their own definitions of student growth for all student populations. States were also able (and strongly encouraged) to include stakeholders in the development of teacher assessment plans.

Although ESSA permitted increased leniency in state defined school accountability plans, several states did not drastically modify their previous plans, citing costs incurred from building the initial value-added systems. However, some states added new non-academic indicators to their accountability plans, such as chronic absenteeism. Furthermore, states also added their own goals for decreasing the achievement gap of underrepresented student populations (Close, et al., 2018). As for teacher evaluation systems, some states shifted in how the evaluation would be used. For instance, some states would allow districts to use the value-added system in formative ways instead of restricting its use to only summative uses (as teachers would use the data to inform their practice instead of being used solely in employment decisions). Other modifications have included local districts being permitted to use their own value-added model and to also include multiple measures of a teacher's success (Close, et al., 2018).

While state directed changes from ESSA have mostly been at the teacher, school and district levels, some states have used the legislation to modify teacher preparation programs. For instance, some states developed programs to increase the diversity of their teacher workforce by incentivizing the recruitment of more diverse candidate into training programs. Additionally, other states used the flexibility in the legislation to modify training requirements by requiring year-long student teaching programs or developing residency programs (which include longer student teaching requirements). Other trends in teacher training have included revised licensure and certification processes, loan forgiveness for teachers who graduate from a training program, as well as the increased use of data systems to both recruit from specific preparation programs and also to monitor program effects (Johnson, 2018). As federal policies have trended toward a focus on state policies and implementation of federal law, the next section will focus on trends in k-12 and teacher preparation throughout the US.

1.3 State Context

Although the federal government has proposed many regulations as it pertains to teacher education, states are ultimately responsible for program approval and review. Indeed, states have a significant incentive to evaluate programs as graduates from state-approved programs can quickly obtain a teaching license to teach in that state (Feurer, Floden, Chudowsky, & Ahn, 2013). As such, states have often sought ways to evaluate programs for the improvement of their state's teaching workforce. Historically, state evaluations have mostly consisted of data on program inputs (such as admission criteria, required course offerings, number of in-service teaching hours, faculty qualifications, and cooperating teacher qualifications) (Henry, et al., 2012). Due to significant data availability limitations, states have not been able to use output metrics beyond teacher career outcomes like license exam scores, rates of hiring, and surveys of graduates and employers (Feurer et al., 2013; Tatto, et al., 2016). However, as more data became available and attention to accountability increased, some states began to use student achievement data and teacher value-added scores to evaluate the success of a program (Bastian, Patterson, & Pan, 2018; Feurer et al., 2013; Goldhaber, 2018).

The first state to adopt a value-added assessment of teacher training was Louisiana. In 1999, Louisiana developed new state standards aimed to improve their teaching workforce within their k-12 schools. The state developed a strategy focused on improved recruitment, hiring, professional development, and training of those wanting to join the profession. The state standards for teacher training required both advanced changes to curriculums (such as increased clinical experiences) as well as changes in reporting requirements. In 2003, Louisiana required all approved programs to use teacher value added data, to evaluate the effectiveness of teacher training programs (Coggshall, Bivona, & Reschly, 2012; Noell, Burns, & Gansel, 2018).

The state created program profile reports that were public and were used internally for programs to evaluate themselves (Noell et al., 2018). States gave each program a report that provided k-12 achievement data for each graduate of their program, which allowed the program to identify patterns amongst all graduates. Programs could then use the data to identify areas of weaknesses and to demonstrate efforts to improve. For instance, using k-12 student scores, some programs noted that their graduates were better able to teach low performing students, but not high performing students. Another program noted that students of their graduates struggled with essay assessments but were strong in grammatical assessments. Programs then examined specific course offerings that teach these concepts as well as concepts that are taught within all courses (such as differentiating learning for diverse student populations). From these assessments, programs added additional courses, resources, and training opportunities in these areas of identified weakness to their teacher training (Feurer et al., 2013). Authors Noell et al., (2018) noted that although these modifications took time, there were improvements across all programs once programs implemented reforms.

Using Louisiana as a model, the states of Tennessee, Florida, and North Carolina soon followed with similar efforts, by linking teacher preparation programs to the value-added scores of their graduates and requiring programs to use that data in program evaluations (Council of Chief State School Officers, 2016). Such data and report cards have since been used to identify programs of lower quality as well as areas where teacher performance was lower. For instance, in the first iteration of their state report, Tennessee identified 12 of their 40 programs as “lower quality” (programs that produce teachers whose students demonstrated lower achievement scores on standardized tests) (Tennessee Higher Education Commission, 2011). However, similar to Louisiana, the state of Tennessee has noted that programs have revised offerings and curriculums based on the data provided to programs. As of 2018, only 4 programs are ranked as lower quality programs (however, some programs originally evaluated in 2010 have ceased offering teacher training) (Tennessee State Board of Education, 2018). In addition to these program specific uses of the data, aggregated reports from all four states were made available publicly, in formats intended to increase information of consumers and prospective teacher candidates (Coggshall et al., 2016). Such efforts have continued throughout states and as of 2016, 16 states use value-added outcome data as one portion of training program evaluations (von Hippel et al., 2016).

Additionally, state reforms of teacher education programs extend beyond using value-added scores in program evaluations. Evidence from a 2017 review of state policies of teacher education by the Council of Chief State School Officers (CCSSO) found that all 50 states (and Washington DC) were in the process of revising state licensure requirements, standards, and required measures of outcome data to be used in training evaluations. While these efforts are present in all states, state reforms are often unique to the state context. For instance, as part of the licensure process, 20 states require that candidates demonstrate mastery of teaching content-specific courses, which states measure through the Praxis Principles of Learning and Teaching exam, edTPA, or a combination of exams (such as passing the edTPA performance exam and a content specific state licensure assessment). Furthermore, an emergent trend has been to require data used in program evaluations and approvals to align with standards used in program accreditation (CCSSO, 2017).

While all states are working to modify data that programs must submit and use in their evaluations, this is one area that is the most diverse and least consistent across states. For instance, 15 states use completer satisfaction surveys at the end of the training in program reviews, whereas 18 use satisfaction surveys from program graduates one year after they graduate. Six states have begun to use satisfaction survey data from multiple constituents (student evaluations of their teachers, the candidate's own review of the program and/or candidate reviews from principals) to review programs. Other data includes: state-based classroom observations of all first-time teachers, teacher retention (within the state and not just in their original placement) as well as placement in high-need schools (CCSSO, 2017). Although the use of data is growing (as are the metrics used), many states are limited in that they do not have longitudinal data systems that are capable of tracking students and teachers through their careers. As a result, many states are limited to programmatic input data until more robust systems are developed in those states (CCSSO, 2017).

1.4 North Carolina Context

While efforts to reform and modify teacher education have been prevalent throughout all states, North Carolina has been a leader in teacher education reforms over the last thirty years. Given North Carolina's robust longitudinal data system, the state has been able to examine the teaching profession noting strengths and weaknesses within their teachers to encourage and influence changes in teacher education. For instance, in the mid 1980's North Carolina, noted both a teacher shortage within their schools and lower enrollment numbers in training programs. In response, the state passed legislation to increase approved alternate teaching routes for those wanting to teach in the state (North Carolina General Assembly, 1985; Henry, Bastian, et al., 2014). Additionally, during this period the state recognized that increasing teachers through these alternative routes did not address (or ensure) a teacher's quality. As such, the state of North Carolina, spearheaded the creation of the National Board for Professional Teaching Standards (NBPTS), throughout the US. This certification is an additional certification that teachers can pursue, and to receive this added credential teachers must demonstrate mastery of teaching techniques and content specific teaching. Teachers are only eligible once they receive state certification (and have taught for more than three years) (National Board for Professional Teaching Standards, 2019). The state used this mechanism as one way to increase teacher quality while increasing the number of teachers, and as of 2018, has the highest number of NBPTS teachers in the US (NBPTS, 2019).

Such efforts to encourage teacher production and teacher quality continued throughout the 1990s when the state created school accountability rating systems (ahead of the enactment of the No Child Left Behind Act in 2001). These efforts sought to reward schools meeting state standards of education and penalize those schools that did not meet standards (Henry, Bastian et al., 2014). As part of this effort, the North Carolina General Assembly required in 1995 that the state build a data system to track student progress and achievement from kindergarten through higher education and into the workforce. This data system has since expanded to produce longitudinal assessments (beginning at pre-k) for school districts, higher education systems, and the state (The State of North Carolina, 2010).

Thanks in part to this robust data system and following the lead of Louisiana and Tennessee, North Carolina was one of the first states to track progress and outcomes of teacher training programs (Henry, Bastian et al., 2014). As part of this effort, the University of North Carolina General Assembly (which oversees the largest teacher training programs in the state) requested an accountability study of teacher training programs within the University of North Carolina system (Henry, Bastian, et al., 2014; The State of North Carolina, 2010). This study examined program effects across grade levels, subject areas, alternative routes to teaching, and effects within student subgroups (Henry, Thompson, Fortner, Zulli, & Kershaw, 2010). Although methodological debates persist regarding this study, authors found that overall, teacher prepared by a UNC institution raised elementary student scores slightly more than those prepared through other routes (including alternative and out-of-state programs). However, as it was commissioned by the UNC system, it only included 15 programs and not all training providers in the state (Coggshall et al., 2012; Henry et al., 2010).

In 2017, the North Carolina General Assembly passed the Excellent Educators for Every Classroom Act (Senate Bill 599/State Law 2017-189) which requires that all programs approved to train teachers in the state to submit reports on their candidates (North Carolina General Assembly, 2017a). Outcome data in these reports include teacher performance measures, metrics used for teacher evaluations, candidate satisfaction surveys, and growth scores of students taught by graduates of their program (North Carolina General Assembly, 2017a). Additionally, the North Carolina Department of Public Instruction provides programs with student unit record data to aid programs in identifying areas to improve. Finally, aggregate reports are publicly available to inform future teachers, districts, and other general audiences of program strengths and weaknesses (North Carolina General Assembly, 2017a). These program performance reports are also part of North Carolina's plan for statewide accountability of teacher training as federally required through ESSA legislation (NCDPI, 2018c).

These changes in teacher education throughout North Carolina are not limited to reporting requirements for training programs, however. For instance, the state modified in-service hours required for all first-time teachers (making in-service training required even for alternative training programs) and once again increased alternative routes and models for teachers to enter the profession and receive state licenses. Lawmakers cited the decline in the number of students in education programs and as such, made entry routes more accessible by allowing non-university programs to offer training options (specifically through on-line options) (North Carolina General Assembly, 2017a). Furthermore, in 2015, citing the literature that holding a master's degree does not impact student achievement, the legislature also removed the financial incentives associated with the added degree. These regulatory and legislative changes have been met with significant concerns from those at teacher training programs. They noted that increased regulations on teacher training and fewer incentives could discourage teachers from pursuing master's degrees, and negatively impact teacher outcomes (even if there wasn't a direct impact on student achievement) (Clotfelter, Ladd & Vigdor, 2010; Sawchuck 2014).

Legislative efforts aimed to improve public education in North Carolina have not been restricted to reforms in data collection and teacher level policies. For instance, in 1996 North Carolina first began to allow charter schools to operate in the state. While the North Carolina State Board of Education and local school districts were involved in both approving the charter school (and subsequently the funding of the school), independent entities could manage and run the school with more autonomy than traditional public schools. Such efforts sought to improve learning for all students across the state, but most specifically for students at risk as they allowed for increased opportunities for innovative teaching methods. Furthermore, charter schools also provided parents the opportunity to have choice in their child's school as well as the educational experience they wanted for their child/children (Ladd, Clotfelter, & Holbein, 2015).

Since charter schools were first established in North Carolina, the number has grown from 33 in 1997-1998 to 196 in 2020 (Ladd, et al., 2015; NCDPI, 2020a). While this is a significant increase over the last 20 years, much of this growth has occurred in the last 8 years, with a 96 percent increase in the number of charter schools in the state since 2011-2012 (and more than double the number of students in that same period taught in charters). As of May 2020, 8 percent of k-12 students (121,231) who were taught in public schools in North Carolina were enrolled in charter schools (and more than half of those students were in elementary school) (NCDPI, 2020a). Student demographics of charter schools mirror those in traditional public schools as slightly more than half of students are white, 26 percent are black, and 11 percent are Latino students. However, researchers have found that recently charter schools have become less racially diverse and that individual schools are racially segregated in that some schools primarily serve racial minorities, whereas others primarily serve white student populations. Researchers also found that while the state itself is racially segregated, student populations are racially segregated by school, wherein white students tend to be enrolled in predominantly white schools regardless of the geographic region of the state (Ladd, et al., 2015). Such regional differences are also seen in the concentration of charter schools, as charter schools are not equally distributed throughout the state. For instance, the percentage of students in charter schools is highest in the northeast region of the state (with 12-27 percent of students in public schools were taught at charter schools), whereas other regions have much smaller percentages of students in public schools (and some regions have no charter schools) (NCDPI, 2019).

Turning to teachers at charter schools, overall 7 percent of all teachers in public schools I North Carolina teach at a charter school (and 53 percent of all charter school teachers were elementary teachers as of May 2020) (NCDPI, 2020b). One difference between charter schools and traditional public schools is teacher requirements. For instance, only 50 percent of teachers in the charter school need to have a license to teach in a charter school, whereas all teachers at a traditional public school must have a teaching license to teach in the school (NCDPI, 2020b). However, those that teach tested grades, such as math and reading must have a degree from a university, though the degree does not have to be in education. Furthermore, national trends show that teachers at charter schools tend to have fewer years of teaching experience (26 percent had less than 4 years), over half only had a bachelor's degree (whereas at traditional public schools 60 percent have a master's degree), and teachers tend to leave after 3 years. Finally, while charter schools have more flexibility in teacher salaries, national trends show that this translates to a decrease by almost $10,000 (with an average of $46,000 at charter schools compared to $55,000 for traditional public school teachers) (Epple, Romano, & Zimmer, 2016).

While charter schools were often designed with the intent to increase both parental choice and student achievement, the research has been mixed on the effects of charters on student learning. As researchers Ladd et al., (2015) found, when charter schools were first introduced in North Carolina, student achievement was higher in traditional public schools, but as charter schools became more established student achievement gains in both reading and math surpassed those in traditional public schools. However, findings show that such differences have recently disappeared and now there are minimal differences in student achievement. This could be because higher achieving students transitioned away from public schools and thus increased achievement in charter schools could be attributed less to the education at charter school and more representative of the students themselves. Other researchers have questioned whether teacher quality at charter schools can be attributed to higher learning gains. For instance, Ozek, Carruthers, & Holden (2018) found that teachers in higher-income charter schools had higher value-added scores, whereas those in lower-income charter schools were not more effective than their traditional public-school counterparts. Findings suggested however that such differences were likely due to increased teacher experience (as teachers with more years of experience tend to be more effective), and less effective teachers were more likely to leave charter schools than those at other public schools. They also found, that charter school teachers tend to be happier in their teaching environment, due to increased salary and more input on their school environment, which may have positive returns to student achievement and teacher value-added scores.

Much research on charter school teachers has examined the movement between traditional public schools and charters. Literature shows that there are strong incentives for teachers to teach in North Carolina, as charter schools have more flexibility in salary structures (which can translate to higher or lower than other public schools), teachers are involved in curriculum design, and often report higher satisfaction in teaching (Epple, et al., 2016). However, research has also shown that as qualifications for teaching at a charter school are lower (such as less experience and not licensed) teachers at charter schools are less effective than other public school teachers (Carruthers, 2012; Epple, et; al., 2016). For instance, as Carruthers (2012) argues, while licensure does not necessarily indicate low effectiveness, that those who left the traditional public school for a charter were lower quality teachers than in other schools. Author Jackson (2012) further argued that switching from a traditional public school to a charter is a predictor of lower teacher effectiveness. Although charter school teachers may be less effective, the presence of charter schools has made it more difficult for traditional schools to staff schools that are already hard to staff by reducing the teaching pool. As a result, traditional public schools have increased salaries to recruit higher performing teachers.

Finally, while the presence of charter schools has had an effect on the teacher labor market throughout states, researchers have found that there is high teacher attrition at charter schools. Indeed, researchers found that teachers at charter schools were more than twice as likely to leave the profession than their public-school counterparts (Cano, Flores, Claeys, & Sass, 2017). Authors argue there are a multitude of reasons which may contribute to attrition. This may be that charter school salaries are lower, there is less training required for teachers, and as teachers are not required to be licensed teachers at charter schools by use teaching as a temporary job until a higher paying non-teaching position becomes available. One of the many reasons this finding is alarming is that high teacher attrition has been shown to predict lower student achievement. As such, authors argue this may be one of the reasons charter schools have not shown significant differences in student achievement when compared to that of traditional public schools (Cano, et al., 2017).

1.5 Local Context

In addition to federal and state reform efforts of teacher training programs, training programs themselves are transforming their training, often spurred by changes from governing bodies. For instance**,** states and programs alike are using edTPA to evaluate teacher training techniques and offerings. Currently, 41 states and DC require all teacher candidates to submit scores of their edTPA evaluations for licensure (AACTE, 2019). edTPA is an assessment tool that uses videotaped evidence of teacher competencies in subject-specific areas and reviews the candidate's ability to effectively teach. Areas that are reviewed include lesson planning, teaching specific lesson modules, and demonstrating knowledge of teaching methods (broadly and subject-specific) (AACTE, 2019; Lewis & Young, 2013). States can then use those scores and tie them back to the training program for program evaluations as well as the candidate's own licensure. While states use scores to review programs, programs can also use this data to see detailed information about an individual candidate's strengths and weaknesses (and as well as specific areas within a candidate's teaching practice). Programs can also aggregate candidate performance as they can see trends of all their candidates as measured by a national rubric (AACTE, 2019; Lewis & Young, 2013). Other local efforts include opportunities to increase practice-based teaching within the training program. Almost all programs have a clinical requirement within the program (with the historical trend being a semester teaching experience inside a classroom). However, research shows that earlier training experiences aid in a teacher's feeling of preparedness once she becomes the teacher of record (even if early experiences are only observational) (Benedict, Holdheide, Brownell, & Marshall Foley, 2016).

Given restraints by locations and the availability of classrooms, programs have had to develop unique opportunities (Benedict et al., 2016). One method used has been for candidates to watch videotaped lessons of teachers and then analyze the teacher's technique (Santagata, Zannoni, & Stigler, 2007). Another option that training programs have added is computer simulations. In these simulations, student 'avatars' serve as practice students for candidates to practice their teaching techniques they learn in their courses before they enter the classroom (Dieker, Straub, Hughes, Hynes, & Hardin, 2013). Additionally, many programs note that limited financial resources can hinder the opportunities available to candidates. As a result, programs have implemented structured tutoring and mentoring programs. For instance, through tutoring and mentoring, candidates are able to apply their instructional practices directly in a supervised environment. These mentors are distinct from the university supervisors in that they provide candidate support but are not typically involved in the assessment of a candidate's practice. For instance, they are alumni of the program, emeritus faculty, or other affiliated program staff (Al Otaiba, Lake, Greulich, Folsom, & Guidry, 2012). Another notable approach is teaching opportunities within the teaching preparation courses themselves. For instance, some programs have implemented microteaching approaches, where candidates plan lessons for their peers, teach the lesson, and then their peers (and professors) provide feedback on the lesson itself (Benedict, et al., 2016).

At the local level in North Carolina, notable reforms are also present. For instance, in 2018, the University of North Carolina System reviewed emerging practices of training programs, as well as survey results from graduates of each program. The goal of this survey was to determine if programs were modifying their practices to match needs of their candidates and graduates. Findings showed that overwhelmingly candidates wanted more clinical experiences. For instance, graduates noted they did not feel prepared to teach diverse student populations (especially those that were not similar to students within their official clinical experience) and they did not adequately know how to apply their training to the classroom (Bryan, et al., 2018).

To address these complaints, programs reported they were modifying clinical experiences in numerous ways. For instance, programs added classroom observations within a candidate's first year, implemented required tutoring, and others added community service with pk-12 students in the region. Other clinical experiences that have been added by some programs include year-long co-teaching opportunities where candidates co-teach a classroom and are supervised by a cooperating teacher (before the candidate has his/her own classroom). One notable method that is being implemented by a program is increased faculty observations of candidate clinical experiences. Candidates noted that cooperating teachers in their area schools were not engaged and as a result, faculty members have taken on a more involved approach to give candidates feedback during their experiences (Bryan, et al., 2018).

Beyond clinical experiences, the UNC System report noted that programs have modified curriculums as well as learning assessments to meet candidate needs. For instance, programs have added more clinical experiences to assist candidate's with teaching a diverse student body as well as additional courses on special student populations, with attention to modifying teaching techniques. Programs have also added behavior management and literacy techniques throughout all courses so that graduates can better understand how to apply these concepts in all elements of their teaching. Finally, program have historically relied on text books and rote assignments and are beginning to incorporate course assignments that include teaching peers, videotaped lessons, as well as independent research to learn effective teaching techniques (Bryan, et al., 2018).

1.6 Research Questions

Given the continued debates at the federal, state, and local levels, it is clear that more research is needed to further inform how we can improve teacher education. To that end, this study aims to fill current gaps in the literature, by understanding what methods within teacher training programs, have an effect on student achievement and a teacher's longevity in the teaching profession. This study will answer the following questions:

1) What program elements, beyond those available in the public domain, are present in teacher preparation programs and are indicators of program quality for elementary teaching programs in North Carolina?
2) What is the relationship between elements of an elementary teacher preparation program and elementary student achievement gains in math and reading in North Carolina?
3) What is the relationship between elements of a teacher preparation program and elementary teacher attrition in the first 8 years of a teacher's career in North Carolina?

There is only one identified study that has previously compiled data on elements within training programs and used those to measure the relationship between student achievement scores (Goldhaber, 2019; Preston, 2017). That study used publicly available data on program elements to test the impacts of those elements on middle and high school teachers within North Carolina public schools. To answer questions one and two, this study will similarly look at program element effects on students in North Carolina (using data from the North Carolina Education Research Data Center— NCERDC). However, this study is different from previous literature as it will combine publicly available data with private program-level data obtained through a survey administered to training programs that show an effect on elementary student achievement gains within the state of North Carolina. Areas that will be collected through the survey are not readily available through the public domain such as faculty training techniques and teacher candidate feedback. This study uses this information on program elements to examine their relationship to student achievement using a two-level multilevel model and a school fixed effects model.

Building off of findings from questions one and two, question three will use the survey findings on elements within teacher preparation programs to test whether there is a noticeable impact on a teacher’s career. Much of the literature on teacher training examines the impact training has on student achievement. However, few of these studies question whether training programs impact a teacher’s attrition and only one has linked elements within a training program to teacher attrition (Ingersoll, Merrill, & May, 2014). As a result, this study will combine teacher attrition data from the NCERDC with program survey results to test this relationship using a discrete hazard function. Goldhaber & Cowan (2014) conducted a similar study in Washington state that examined the impact of teacher education on teacher attrition, however that study did not attempt to isolate program elements to test whether the inclusion of certain elements and characteristics influenced teacher careers. Instead, Goldhaber & Cowan (2014) solely examined whether teachers from specific programs were more likely to stay in the teaching profession in the 10 years after receiving their certification. To fill this gap in the research, I will examine the effect of program elements at private and public universities in North Carolina on elementary teacher attrition within the first eight years of their career.

This chapter has provided a current context for this study, both through the research on teacher education but also federal, state, and local concerns that have influenced the design (and reform) of teacher education. The following chapters in this study examine the previous literature on teacher quality and teacher education, and then follows with study methods, results, and ends with implications of study findings. Chapter two serves as the literature review and examines what is known about teacher quality, and specifically, the qualities of a teacher and a teacher's training that have been shown to impact student achievement. The next chapter, Chapter 3 walks through the details of the study, outlining the survey design process, the survey itself, the student and teacher data available for the study, and the statistical models that will be employed to answer research questions two and three. Chapters 4 and 5 provide a detailed explanation of study results. Finally, chapter 6 discusses research findings in the context of prior literature, details what findings mean for programs and the districts that hire program graduates, as well as potential future direction in research on teacher training elements and training programs.

2.0 Literature Review

Throughout the literature on student achievement and learning, one thing has wide consensus amongst scholars: a teacher is the single-most important in-school factor to impact student achievement (specifically as measured by student scores on standardized tests) (Rice, 2003). Researchers have found that the most effective teachers can increase a student's achievement gains by1.5 years, while students with similar characteristics and demographics who are taught by less effective teachers have shown only half a year of increased achievement (Hanushek, 1992; Hanushek & Rivkin, 2006). Further research into a teacher's effect, shows that teacher quality explains 12-14 percent of student gains in mathematics and seven percent of gains in reading (at the elementary level) (Nye, Konstantopoulos, & Hedges, 2004). Translated in financial gains, increasing a student's achievement by one standard deviation can increase a student's earnings by 10-20 percent in his/her lifetime (Chetty, Friedman, & Rockoff, 2011). Researchers have further argued that if less effective teachers were replaced with higher (or average) quality teachers, a student's lifetime earnings could increase by approximately $25,000 and between $250,000-$400,000 for an average classroom (Chetty, Friedman, Hilger, et al., 2011; Hanushek, 2011). This literature provides strong evidence that teacher quality can affect student achievement scores. However, beyond a broad consensus that teachers have an effect, there is less agreement why one teacher can increase a student's achievement more than another.

To better understand why an individual teacher has different effects on a student's achievement than her colleagues, researchers have examined 'qualities' of individual teachers and her teaching practice instead of overall teacher quality. As author Kennedy (2008) notes, teachers must perform multiple, simultaneous, and, at times, competing competencies such as managing behavior, teaching dynamic content, engaging student participation, expanding student thinking, and accommodating individual student needs (Kennedy, 2106). Kennedy further argues that using the term teacher *qualities* emphasizes “there are many dimensions of teacher quality and if we want to improve the broad, general quality of the American teacher workforce, we need to sort out all the specific qualities that are part of it” (Kennedy, 2008, p. 60). Because teaching and the teacher herself comprise so many dimensions, moving to an exploration of these elements can provide a clearer understanding of qualities of individual teachers and teaching that may, or may not, ultimately effect a student’s achievement.

Because of the dynamic nature of teaching, and the multiple attributes that comprise individual teachers, this chapter will explore the relevant literature on a teacher’s individual characteristics and teaching qualities that may affect student achievement gains (as measured by standardized tests). While there is ample literature on a teacher’s effects on elements of student behavior, socialization, civic identity, and beliefs (among others), this review is limited to a teacher’s (and his/her teaching) qualities that may influence student scores on standardized exams. The literature included in this chapter is divided into three sections: attributes that can influence a teacher’s quality, differences between teacher training programs, and finally measurable elements within teacher training programs. The first section of this chapter reviews the research on teacher qualities, and specifically examines those elements that have been found to have a significant (or not) relationship to student achievement. While there are vast differences between individual teachers not all are measurable on a broad scale, as a result, this review will focus solely on observable traits such as a teacher’s academic achievement, the level of content knowledge in the subject taught, and years of experience teaching. This review then turns to research that uses state-level longitudinal data to link student achievement, teacher-level data, and teacher training programs to examine the effects of specific training programs on teachers and subsequently student achievement scores. While findings from this body of literature on teacher training has grown over the last decade as data systems become more robust, the literature on specific attributes within a training program and their effect on teacher quality is limited (Goldhaber, Liddle, & Theobald, 2013). As a result, the last section of this chapter will end with the scant literature on aspects of teacher training that have (or have not) demonstrated an impact on student achievement such as content course, pedagogy, and student teaching.

2.1 Teacher Qualities and Student Achievement

2.1.1 Teacher Certification and Licensure

The first aspect of a teacher to be examined in this review is the impact of teacher's type of certification or license[4]. Teachers, in all 50 states are required to have state issued licenses and/or official certification to teach in k-12 schools. Arguably, this has been to regulate the profession to prevent unqualified teachers from teaching. Yet, as the demand for teachers grew, so did the need to provide more avenues to teaching and modify teacher qualifications and as a result, different multiple certifications grew (Clotfelter, Ladd, & Vigdor, 2010). As certification requirements became more attainable to persons with varied backgrounds and qualifications,

[4] Unless otherwise noted, certification and licensure will be used interchangeably, referring to the type of state requirement the teacher holds to be permitted to teach in a k-12 classroom (such as standard, alternative, emergency, or those required to teach in private schools).

researchers have examined whether this reduced the quality of teachers or, if the profession has been strengthened through an influx of diverse teacher backgrounds and qualities (Sass, 2011). However, research to date has been inconclusive and there is scant evidence that a teacher's certification type is an indicator of her/his effectiveness (Sass, 2011).

In 2000, researchers Goldhaber& Brewer examined a nationally representative sample of students (National Education Longitudinal Study of 1988) to test the effects of teachers with standard and non-standard certification (such as emergency, private, temporary, or no certification) on their student's achievement. Findings showed that students of teachers with standard certification in mathematics performed better on standardized exams than those with teachers who held private or no certification in that subject (with approximately a 1.3point increase in achievement scores). However, they also found that there were no significant differences between students of teachers with standard certification and those with emergency certification (Goldhaber and Brewer, 2000). Such findings set off a larger debate in the field as researchers debated whether a teacher's certification ultimately had an effect on student achievement. Darling-Hammond, Berry, & Thoreson (2001) disagreed with Goldhaber and Brewer (2000) and argued that teacher certification and training does matter in regards to teacher quality, and that the only difference between teachers in the 2000 study was the certification type as some had emergency credentials and other held traditional certifications was their certification (Darling-Hammond et al., 2001; Wayne & Youngs, 2003). They further argued that teachers with more experience and standard educational training (and training in education specifically) have a greater effect on a student's achievement than those with less experience and who did not hold standard certification (at the $p<0.05$ for math and $p<0.001$ in science). However, both research teams noted that that assessing the effects of teacher certification on a student's achievement was challenging given the different state definitions and requirements for teacher certification (Wayne and Youngs, 2003). As a result, they encouraged further analysis at the state level where teaching requirements and certification types are aligned.

In 2007, Clotfelter, Ladd, & Vigdor furthered this debate with research data on teacher credentials and student achievement in the state of North Carolina. They found that teachers with emergency or provisional licenses had a significant negative effect on student achievement (ranging between -0.033 and -0.059 at the $p<0.01$), however teacher experience was one of the biggest indicators for gains in student achievement. Sass (2011), further explored this question on the effects of certification using Florida statewide data that linked student scores on achievement exams to their teachers (and included teacher qualification data). Noting the state context that Florida had multiple alternative routes to teaching, he was able to more easily examine alternative certification effects on teaching. His findings showed that teachers certified through certain alternative methods (such as the ABCTE program which requires only the passage of a certification exam and no coursework) performed better than those with traditional certifications (by 0.06 and 0.11 of a standard deviation in student learning). Sass (2011) argued these findings showed that teacher effectiveness was determined more by a teacher's intellectual ability or experience teaching than certification type. While much of this literature on effects of teacher certification has been inconclusive with marginal differences between types of certification/licensure most findings have agreed with Sass (2011) and suggested that certification is one of the least impactful measures of a teacher's effectiveness to raise student achievement (as measured by student test scores) (Boonen, Van Damme, & Onghena, 2014).

2.1.2 National Board Certification

Although the literature on teacher certification and licensure is mixed, some have argued that additional certification has been found to have positive relationships on student achievement. In particular, the National Board for Professional Teaching Standards (NBPTS) certification, which is an additional voluntary certification for those with three years of teaching experience (NBPTS, 2018). The NBPTS certification is of particular importance to this review as North Carolina has the highest percentage of NBPTS certified teachers in the United States (18% of all NBCT's are in North Carolina) (Belson & Husted, 2015; NBPTS, 2018). One possibility for this is that teachers in North Carolina receive one of the highest financial incentives of any teacher in the US for those that receive National Board Certification. As of 2018, the North Carolina Department of Public Instruction provides financial support for teachers to pursue the certification and a 12 percent pay increase for teachers in public schools who receive National Board certification (NBPTS, 2018).

Although there are strong incentives for teachers to obtain this certification in North Carolina, the research on its effects on student achievement are mixed. For instance, researchers Sanders, Ashton, & Wright (2005) tested certification effects using student test scores on state exams in North Carolina from 1999-2003 and results were inconclusive as there was no effect compared to those without the certification. In a following study, Goldhaber & Anthony (2007) argued that NBPTS certification served as a signal of an effective teacher for elementary students in North Carolina as the process requires reflection of one's teaching practice. Their findings showed that those who eventually obtained the certification were generally more effective than those that did not obtain the certification (but attempted the process) and those that never pursued it (with a gain of 0.04 of a standard deviation for reading proficiency and 0.05 of a standard deviation in math). However, they also found that it was not the certification that separated effective teachers from ineffective teachers. Instead, they argued that teachers who pursued certification had higher student achievement gains than other teachers prior to receiving certification, and that the certification process does not show any improvement in the teacher's effectiveness. Following this study, Clotfelter, et al.,(2007) found similarly positive results for NBCTs on student achievement when compared to their non-Board certified colleagues. Specifically, they found that that teachers who pursued Board certification had higher student achievement gains prior to receiving certification (with an effect of 0.061 for math and 0.026 for reading one year before a teacher achieves certification). Authors thus argue that Board certification serves as a signal of highly effective teachers, but not that the certification itself influences teacher effectiveness.

Researchers Harris & Sass (2009) explored similar questions using student achievement records linked to teacher certification and credentials in the state of Florida for elementary and middle school students. In their analysis they argued that unlike the Goldhaber & Anthony (2007) analysis, NBCTs are not randomly assigned to classrooms, and that, their models exclude student and school fixed effects, leading to an overestimate of differences between a teacher with National Board certification and those without. When fixed effects were added, Harris and Sass (2009) did not find evidence that the certification process improves a teacher's effectiveness. Their findings showed small differences between middle school math teachers who eventually received certification and those who did not (and is relatively small with an effect size of 0.06 at the 0.05 level). This difference goes away once those teachers received certification.

In 2016 researchers Cowan & Goldhaber, studied both the effectiveness of National Board Certified teachers to raise student achievement, and tested whether a teacher's score on the certification exam is an indicator of effectiveness (regardless if the teacher eventually achieves certification). They found that middle school math teachers had the greatest difference between board certified and non-board-certified teachers with gains of 0.05 standard deviations for students of NBCTs. However, authors noted there was no significant difference between a teacher that initially failed the certification exam (but eventually obtained the certification) and those without the certification. From the literature reviewed here, differences in student achievement of a teacher with Board certification and those without are marginal and depend on a teacher's subject area (with effects being primarily in mathematics). However, the evidence does suggest that interest in pursuing Board certification is a strong signal of teacher quality, though obtaining certification itself is not a strong indicator (nor is there evidence that the process of obtaining Board certification increases student achievement). Researchers have argued that there are other important effects from Board certification. For instance, Belson & Husted (2015) argue that there are spillover effects from certification, and that a teacher with Board certification can positively influence other teacher's performance within her school through mentoring and demonstrating differentiated teaching practice. Furthermore, they provide limited evidence that teachers with Board certification are more likely to pursue leadership roles (thus furthering the spillover effects of quality teaching).

2.1.3 Alternative Certification: Teach For America

Although the research on certification (National Board or other) is mixed with limited significant findings, many experts have argued that certification type is less important than a teacher's training and often reference the literature and findings from studies on Teach for America (TFA). Teach For America is a highly selective alternative certification route to teaching that places members in high-need[5] classrooms for two-years (TFA, 2018). TFA is contextually relevant in North Carolina, as research on TFA in the state of North Carolina shows that TFA is one of the most effective teacher education program in the state of North Carolina (when measured by student achievement on state assessments) (TFA, 2018).

The literature on TFA's effectiveness is widespread, with more than forty-five different studies on the effects of teachers from TFA on student achievement (US Department of Education, 2016b). While the literature is somewhat mixed on effects in other subject, a review of studies on the effectiveness of TFA from the What Works Clearinghouse within the US Department of Education (2016b), found that overall, teachers who were trained through TFA had a large and positive impact on student achievement in mathematics. One of the earliest studies to find evidence of TFA's impact on mathematics was in 2006, when authors Glazerman, Mayer, & Decker implemented a randomized control study in 17 different elementary schools across the US, examining student test scores on the Iowa Basic Skills Test. In their comparisons between student scores for students taught by TFA members and non-TFA trained teachers, researchers found a statistically significant effect on math scores (0.15 standard deviations at the $p<0.01$ level) but no significant findings for English/language arts scores. Using linked student achievement and teacher credential data from New York City schools, researchers Boyd, Grossman, Lankford, Loeb, & Wyckoff (2006), found that teachers entering the profession through alternative routes (such as Teach For America) had higher gains than those entering with emergency

[5] High-need classrooms are defined in this context as classrooms within public schools (in urban or rural locations) that have high percentages of low-income students or otherwise disadvantaged students where teacher shortages are prominent.

credentials/certifications (though not in all subjects). For instance, they found that students of teachers with traditional or alternative certification (specifically TFA) had higher gains in mathematics than those taught by teachers with emergency licenses (with an increase of 3.3 standard deviations). Researchers Kane, Rockoff, & Staiger (2006) furthered this analysis by using a similar dataset from New York City schools but added an additional year and admission data to their sample. Findings showed marginal increases in student achievement scores in mathematics for those certified through Teach For America (a 0.02 standard deviation increase compared to those with standard or other (emergency) certification). Even smaller differences in student achievement scores in reading were found for those taught by teachers with standard certification (increases by 0.01 standard deviations).

In 2011 Xu, Hannaway, & Taylor used scores from end of grade statewide exams in North Carolina, to analyze the effects of TFA members on student achievement in science, reading, and mathematics. Findings were not significant for reading or math but did show a positive impact on student achievement in science (with an effect size of 0.19 at the $p<0.05$ level). In a study in 2012, Turner, Goodman, Adachi, Brite, & Decker examined effects of TFA members on middle school student math achievement using standardized test scores from the Texas Assessment of Knowledge and Skills (TAKS). Findings showed an increase of 17 points on the 2010-11 TAKS mathematics exam for students taught by TFA members (an effect size of 0.l9). Turner et al. (2012) also found a significant increase in student learning in mathematics for students of TFA alumni (i.e. those trained and certified by TFA but who had finished their two-year commitment), with an increase of 23 points in exam scores (and an effect size of 0.27). In their analysis, they also found significant increases in student reading scores (scores were 10 points higher with an effect of 0.11), however this was not evident in their analysis on scores of students with teachers who were currently members of TFA. Following this study, Clark et al. (2013), examined middle and high school student achievement in math using state exam scores for middle school students and a study-specific exam for high school students. The study included student test scores from eight states and 45 schools. Similar to previous studies, findings showed an increase in math achievement by 3 percentile points, with an effect size of 0.07 at the 0.01 level.

Authors Henry et al. (2014) continued this analysis of the effects of TFA members on elementary, middle, and high school students, using end-of-grade exam scores in North Carolina for mathematics, reading, and social studies. This study examined effects of training routes to teaching (such as traditional, teaching fellow programs and TFA). Researchers found that TFA members were more effective than traditionally trained teachers or those with more experience in mathematics for all three grades level examined (with an effect size ranging between 0.07-0.20 at the 0.01 level). However, there was no significant finding for reading or social studies. In 2015, Clark, Isenberg, Liu, Makowsky, & Zukiewicz examined the effect of TFA members and non-TFA teachers on student elementary mathematics and reading test scores (using end-of year tests of students in grades pk-5th). This study continued the research from Clark et al. (2013) by adding states and schools to the original analysis. TFA members were in their first or second year of teaching, whereas comparison teachers had on average 13 years of experience. Again, findings found significant effects in reading for students in grades pk-2nd (with an effect size of 0.12) as well as math for students in grades 1-2nd (with an effect size of 0.16, and the 0.05 level). Although no discernible effects were noted for students in higher grades, authors note that for higher elementary grades (3-5), TFA members were as effective as traditionally trained teachers with more experience. One takeaway from the literature on the effects of TFA is that although the teachers within the program may be more effective at raising student achievement scores (at least in mathematics), there is a limited understanding of what makes members more effective.

Researchers have argued higher achievement scores of students taught by TFA members could be because TFA teachers are from more selective institutions (thus assuming these teachers could be academically high achieving themselves) (Henry, Bastian, & Smith, 2012). However, while there is some evidence that TFA teachers have a positive effect on student achievement gains in mathematics, researchers caution that those who enter the profession through selective methods (such as TFA) are less likely to stay in the profession and report higher dissatisfaction with teaching. This can be a result of the nature of the profession (lower pay and growth in career opportunities) and also lower expectations entering through selective routes to continue in the profession (Kelly & Northrop, 2015). This is particularly relevant to studies on student achievement, as teacher effectiveness has been found to increase with years of experience, and thus even if teachers that enter through selective methods are marginally more effective on the onset, by leaving the profession early (within their first three years) student achievement gains will be minimal (Ronfeldt, Lankford, Loeb, & Wyckoff, 2011).

2.1.4 Teacher Qualifications: Master's Degree

Throughout the literature on teacher qualities and their effects on student achievement, one of the credentials that has received significant attention from researchers is the impact of (or lack thereof) a master's degree. The literature overwhelmingly suggests that teachers with master's degrees are no more effective at raising student achievement scores on standardized exams than teachers without master's degrees. For instance, Aaronson, Barrow, & Sander (2007) sought to examine the relationship between teacher attributes (such as experience and credentials) and their effect on student achievement. Using panel data from elementary student test scores in Chicago public schools, researchers argued that while differences in student achievement could be attributed to individual teachers, a teacher with an advanced degree was no more effective than one without and saw no significant difference between the two teachers. In a similar study of student achievement in North Carolina, authors Clotfelter et al. (2007) examined 10 years of teacher data (1990-2000) that was linked to elementary and middle school student test scores in reading and math. This study also found no significant effects (and at times a negative effect) on student achievement from having a teacher with graduate degree. This finding was further confirmed in a study using longitudinal panel data on elementary student test scores in San Diego and Los Angeles (Buddin & Zamarro, 2009).

Although findings have not shown effects of teachers with master's degrees on student achievement, until recently, several states determined teacher salary levels according to the teacher's educational attainment. However, citing the lack of evidence, the state of North Carolina, discontinued the pay incentive for increased education in 2013 (Sawchuck, 2013a; Chingos, 2014). Because of this change, authors Ladd and Sorensen (2015) reexamined the literature, to adjust specifications in previous analyses with an updated dataset (using student and administrative records from the state of North Carolina between the years 2006-2013). In their analysis they used fixed effects to control for time-invariant characteristics of teachers. Their rationale was that they would be able to separate teacher decision effects from the effects of having a master's degree, arguing that teachers choose to pursue (and ultimately receive) a master's degree. However, even with these alterations, findings showed no significant effects on student achievement from a teacher with a master's degree. When they applied teacher choice fixed effects, findings were not statistically significant, nor were findings significant when they take out the fixed effect model.

They did find effects that were not tied to student achievement as there was a marginal significance in student attendance for students of teachers with a master's degree. For instance, a master's degree in school administration had a 7.7 percent increase on student attendance (and the $p<0.01$ level) and master's degrees in math had a 2.3 percent increase in student attending (in middle and high school grades) (Ladd and Sorensen, 2015). Although no other discernible effects on student achievement have been tied to teachers with master's degrees, several states still provide financial incentives for master's degrees. One reason for this is that states and school districts see benefits that are less measurable but still valuable in a teacher, such as motivation, higher academic ability, pursuit of school leadership roles, and dedication to the profession (Chingos, 2014; Ladd & Sorensen, 2017).

2.1.5 Years of Experience

One teacher attribute that has wide consensus among researchers and has consistently shown an effect on a student's academic achievement is the number of years the teacher has taught in a classroom. However, there is limited agreement as to the number of years a teacher needs to teach to reach her full effectiveness, or if a teacher's effectiveness stops, or even decays after a certain number of years in the classroom. For instance, research using panel data from North Carolina and Florida showed that teachers with one year of experience are more effective at raising student achievement than teachers with no years of experience, however the impact of experience begins to wear off after the teacher is in the classroom for five years (Clotfelter, et al., 2007, and Henry and Sass, 2007). Clotfelter et al. (2007) noted that for elementary and middle school teachers, experience increased student achievement with each year of teaching, with the highest gain of 0.119 standard deviations for mathematics and smaller, yet significant, findings for reading

(0.096) at the $p<0.01$ level, however these gains slowed after five years. Similarly, authors Croninger, Rice, Rathbun, & Nishio (2007) tested the effects of teacher experience using a national longitudinal survey of student achievement (the ECLS-K) for students between the years of 1998-1999. They found that students taught by teachers with two or less years of experience had lower gains on reading exams than students taught by teachers with more years of experience (with a -0.06 decrease of a standard deviation in learning at the $p<0.1$ level). No significant findings were found for math achievement scores or for students taught by teachers with five or more years of teaching.

Authors Boyd, Lankford, Loeb, Rockoff, & Wyckoff (2008) were one of the first to unpack the effects of teacher experience on student math achievement by using achievement scores in New York City to test when (if ever) gains from teacher experience begin to diminish. They found that gains in math for grades 4-8th saw the largest increases from teachers in their first year (with an increase of 0.06 standard deviations for grades 4-5 and 0.04 standard deviations for grades 6-8). However, they also found that teachers with 20 or more years of experience were no more effective at raising student achievement than teachers with only five years of experience. Additionally, Ladd (2008) examined North Carolina panel data on high school student math achievement and found that teachers with 25 or more years of experience were less effective than those with less experience. This suggests not only that gains do not increase after teachers are in the classroom for five years, but also that they begin to decline and lose effectiveness after 20-25 years of experience teaching.

Although early researchers saw marginal returns to student achievement from teachers with five or more years of experience (and a decay after 20 or more years), recent debates have found contrasting results, thus continuing the debate as to whether a teacher’s effectiveness diminishes (and if so, when). For instance, Chingos and Peterson (2010) found modest gains in effects in middle school student math achievement from teachers with 20 or more years of experience. Authors noted that while gains at 20 years were modest, gains were still present. Further, Papay and Kraft (2015) used elementary and middle school data in mathematics and reading to test different model specifications to see if years of experience saw a decay in student achievement. While model specifications did produce differing results, they found that overall, in every model, gains were still seen after five years, continuing to year 29. Conversely, Henry, Bastian, & Fortner's (2011) work using North Carolina panel data of student test scores for middle and high school students found that while teacher gains increased during the first two years of teaching, there were no significant effects on student achievement after the third year of teaching. They do note that teachers that stay in the profession for five years, were more effective than those that leave the profession (arguing that those that stay are more effective than those that leave), yet they found no noticeable differences between year five to year 20.

While studies have consistently found that teachers were more effective with more years of experience, researchers recently noticed that positive effects were only found when teacher fixed effects were added to the statistical model (Kini & Podolsky, 2016). In a review of 30 studies on teacher experience and student achievement, Kini & Podolsky (2016) found that of the 18 studies which applied teacher fixed effects to their statistical model, these models found positive gains in student learning as teacher experience increased regardless of subject or student level. However, of the other 12 studies reviewed that did not apply teacher fixed effects to their analysis only two found mixed results (while the others found no significant findings). Authors suggest adding teacher fixed effects allows researchers to compare a teacher's current effectiveness to that of their previous effectiveness when the teacher had less classroom experience, and thus is the preferred model specification for this research.

Noting the differences in model specifications, Wiswall (2013) sought to test findings from earlier studies that found positive gains in student achievement from teacher experience when statistical models were modified to include teacher effects. Wiswall (2013) used fifth grade student and administrative panel data from North Carolina to test findings from Clotfelter et al. (2007) as well as those of Harris & Sass (2011). This study applied teacher, student, school, year, and classroom fixed effects to model specifications used in previous studies to test the robustness of findings of teacher effects. The author replicated findings in fixed effect models (showing years of experience had an effect on student achievement in math and reading) but did not find gains when using OLS and random effects models. Such sensitivity to model specification demonstrates the sensitivity of the findings in this research and can provide context for the hesitation within the research community to state anything more than teacher experience can have a positive impact on student achievement.

2.1.6 Subject Matter Expertise

While teacher certification and degree levels (such as master's degrees) have not consistently shown significant effects on student achievement, researchers have found evidence that subject matter expertise can impact student outcomes. For instance, authors Croninger et al., (2007) examined teacher characteristics and their effects on student achievement using data from the Early Childhood Longitudinal Study, from the US Department of Education. In this study, they found that elementary students taught by teachers with training in elementary education saw higher gains in reading (by 0.08 standard deviations) than those taught by teachers with training in other subjects. They did not however, find any differences in mathematics achievement and argued this could be due an emphasis on student reading in elementary education training programs. In a study of elementary student achievement data from Florida, authors Harris & Sass (2007) found that educators with majors in general education were less productive than those with a non-education major, yet English/language arts teachers who majored in ELA education were found to be more effective. Clotfelter et al. (2007), did a similar analysis of the subject-knowledge of teachers using data from North Carolina test scores of high school students and teacher scores on subject exams. For instance, researchers found that a teacher’s score on the math Praxis exam translates to an increase of 0.03 standard deviations on student achievement in algebra exams. Clotfelter et al.(2007) found that students taught by teachers without scores in the subject that they taught (suggesting the teacher did not have enough training in the subject to take the exam), had lower achievement exam scores.

In a study of student achievement in New York City, authors Rockoff, Jacob, Kane, & Staiger (2011), administered surveys to teachers that tested (among many items) a teacher’s content knowledge in math using the Mathematical Content Knowledge Test. They found that teacher content knowledge in math was a stronger predictor of student achievement on math exams (with a coefficient of 0.03 at the $p<0.02$ t level). Furthermore, they found that student growth scores in math was higher for those taught by teachers who majored in math (0.04 standard deviations at the $p<0.02$ level). While, most studies of a teacher’s content knowledge use teacher scores on licensure exams prior to entering the profession, authors Garet et al. (2016) examined effects of professional development to increase a teacher’s reading and mathematical knowledge (and subsequent impact on his/her student’s learning in those subjects). They found limited, if any, significant effects on student learning, even though teacher content knowledge did improve. For instance, teachers who received the content focused professional development scored 21 percentile points higher than teachers without the treatment (indicating higher content knowledge), yet students of those teachers scored two percentile points lower than students of teachers that did not receive the treatment.

In another study that used a national dataset from Peru, Metzler & Woessman (2012) were able to test teacher content knowledge in math and reading and link those scores to sixth grade student scores on national exams. In this study, researchers found that one standard deviation in teacher's math scores raises a student's math achievement by 0.09 standard deviations (no significant findings were observed for student reading). Authors also noted that content knowledge (and thus a teacher's ability to do well on an exam of the subject he/she teaches) could also be tied to a teacher's cognitive ability, suggesting a teacher's intellectual ability could be the driver of higher student test scores (Metzler & Woessman, 2012; Coenen, Cornelisz, Groot, Maassen, & Van Klaveren, 2018).

Another study of the effects of a teacher's mathematical content knowledge on student achievement, Hill, Kapitula, & Umland (2011) argued that increases in student achievement was not restricted to those with just strong mathematical content knowledge, but also teachers who could recognize student ability and mathematical learning. They examined what factors influenced student achievement scores, as measured by student scores from standardized exams as well as classroom observations and interviews of middle school teachers. Researchers found that comprehension of math content was partly correlated to student achievement, but more so, a teacher's mathematical knowledge was strongly correlated to her ability to teach mathematics (correlation of 0.90 as measured by the in-classroom instruction instrument and the survey instrument correlation was 0.72, both at the $p<.01$ level). Thus, they argue that just having a strong understanding of mathematical content is not what would increase student achievement, but more it was because they had strong content knowledge that they were better able to teach difficult concepts, and that influenced student achievement. Campbell et al. (2014) found similar findings that a teacher's mathematical content knowledge increased mathematical achievement scores for upper-elementary students by 7.1 percent (at the .033 level) and 16.6 percent (at the $p<0.001$ level) for students in the middle grades. Additionally, authors noted that when they controlled for student demographics and prior achievement, a teacher's mathematical content knowledge had a significant effect on upper-elementary student achievement for those teachers that had an awareness of student ability (effect size 0.054 at the $p<0.014$ level).

2.1.7 Teacher Achievement and Entry Exams

Much of the debate and research on the effects of teacher qualities on student achievement examines training, or more what can be done to "build a quality teacher". However, one alternative argument, is that quality is determinant on the individual teacher herself and her own abilities, such as individual academic and intellectual abilities. In this argument, low teacher quality could be the result of ineffective recruitment as those who are recruited and enter into the profession have low academic abilities. Indeed, widespread evidence has shown that academic abilities of teachers are low compared to other professions (Henry, Bastian, & Smith, 2012; Lankford, Loeb, McEachin, Miller, & Wyckoff, 2014). Researchers and policymakers have noted this decline and suggest that a more selective teaching workforce could translate to higher student achievement scores. For instance, the literature on teachers in the TFA and those in teaching fellow programs has shown higher student gains, and entry requirements are more selective than for those that enter the profession through traditional routes (Henry, Bastian, & Smith, 2012). One theory for this increase in student achievement is that those who achieve higher academically tend to have a stronger command of their subject (such as math) resulting in higher teaching quality and ability to teach the subject to others. However, a contrasting argument can be made that a high achieving teacher herself may not be able to teach (or reach) those that struggle academically (Kelly & Caughlan, 2011). For instance, evidence from Jennings & Diprete (2010) show that teachers who have the highest impact on student achievement do not always have the highest impact on social and behavioral development. Thus, those that may be high achieving may not have the skills necessary to address the many dimensions of teaching such as socialization and/or differentiated student learning.

To test the influence of a more selective teaching workforce, researchers have examined whether observable teacher qualities tied to a teacher's intellect such as teacher's GPA, score on entry exams, as well as licensure exam scores measure whether teacher ability can influence student achievement (Henry, Bastian, & Smith, 2012). Author Goldhaber (2007) examined relationships between teacher licensure exam scores (both on curriculum and content Praxis II exams in math and reading) and student achievement using data for students in grade 3-6 in North Carolina public schools. In this analysis, he found that teacher scores had a small, but significant effect on student achievement. For teachers that scored in the top quintile on the reading curriculum test, student achievement (on statewide exams) was 2.2 percent of a standard deviation higher than those of students taught by teachers with lower scores. For math, this was even higher as students scored 3.5 percent of a standard deviation above students of teachers with low curriculum scores (both at the $p<0.01$ level). Findings from teacher content exams were only significant on the math exam, with a 2.4 percent increase in a standard deviation on student scores (at the $p<0.05$ level), for students of teachers that scored in the top quintile.

Using longitudinal data for ten years of data from North Carolina, Clotfelter et al. (2007) examined similar findings for students in grades 3-5. This study had similar findings to Goldhaber (2007) in that they found that students of teachers with test scores that were 2 standard deviations above the mean had an increase in their math scores by 0.068 standard deviations. Similarly, for students of teachers with scores that were 2 standard deviations below the mean, math scores decreased by 0.062 standard deviations (both at the 0.01 level). Findings on student reading achievement exams were significant for teachers that scored the highest, but no significant findings were present for students of teachers with lower scores (with gains of 0.002 standard deviations at the $p<0.05$ level). In a follow-up study, Clotfelter, et al. (2010) attempted to resolve the concerns that non-random sorting of students into classrooms may cause on results. In their 2010 analysis, they applied student fixed effects to their model and removed time dimensions. They found that teacher test scores were predictive of student achievement, however smaller than in their previous study as a gain of one standard deviation in a teacher's math score translated to an increase of 0.047 standard deviations in student achievement.

Boyd, Lankford, Loeb, Rockoff, &Wyckoff (2008) examined gains in student mathematical achievement for 4th and 5th grade students in New York City public schools. In this study, authors noted that NYC public schools had changed recruitment practices of teachers by hiring teachers with higher scores on the SAT and LAST (teacher certification exams), as well as hiring teachers with more experience and other indicators of higher quality. When authors isolated effects of individual teacher characteristics, they noted that teacher scores on SAT exams were related to higher student achievement scores, as one standard deviation increase in a teacher's math SAT score saw an increase in student achievement by 0.04 standard deviations. Authors also argued that stronger effects occur if all teacher qualifications are combined (such as higher scores on SAT, LAST, competitive college education, and more experience). Findings also showed that the increase in teacher qualifications increased student achievement by 0.03 standard deviations especially for students in the highest poverty schools (which had previously seen the lowest gains in achievement), subsequently reducing the gap in achievement by 25 percent. When experience was dropped from the model and just teacher achievement was added, there was a 0.02 increase in standard deviations in student achievement.

Authors Kane, Rockoff, & Staiger (2008) also examined student achievement data in New York City schools, and explored the differences in teacher qualifications, as defined by his/her certification level (using alternative certifications such as the NYC Teaching Fellows program and TFA as a marker for higher academic). Findings showed limited, if any, effects on student achievement from teachers with different credentials. In their baseline models, they found students of teachers in the NYC Teaching Fellows and TFA programs had achievement scores below traditionally certified teachers (with a decrease of 0.20 and 0.28 standard deviations, respectively). When authors added a student's prior year performance to the model, these differences disappear (although students of TFA members had a marginal increase of a 0.01 standard deviation). Furthermore, when researchers isolated the effects of NYC Teaching Fellows GPA's and SAT scores, no significant relationship was found between a teacher's academic credentials and student achievement.

Using student achievement and teacher administrative data for the state of Florida, Harris & Sass (2008) sought to isolate the effects of teacher characteristics on student achievement scores using state-wide math and reading exams for grades 3-10. One teacher variable they examined was teacher SAT scores and their relationship to student achievement scores. In this analysis, they found no significant relationship between a teacher's score (whether the highest score or lowest score) and student scores. Chingos & Peterson (2010) used similar data from Florida for grades 4-8, only this time they examined the effects of failing a state teacher certification exams and student achievement scores for math and reading. In their analysis, they found that a teacher who failed her/his certification exam had a significant negative impact on student achievement (0.04 for 4th grade math, 0.03 for 6th grade math, and 0.03 for 6th grade reading all at the $p<0.01$ level; there was no significant effect on 4-5th grade reading). While this was a significant finding, authors note, only two percent of the teaching population in Florida failed the certification exam, thus suggesting the findings are less relevant.

In two international studies, teacher cognitive ability was found to have a limited, if any, influence on a student's achievement. Gronqvist & Vlachos (2008) examined teacher administrative data linked to student achievement data in Sweden, and test for effects of teacher cognitive and non-cognitive (social) abilities. Cognitive and non-cognitive skills are based off of required military exams for all males in Sweden (used in determining draft eligibility). This data included GPA, as well as exams used to test a person's skill in interacting with others, problem solving ability, and ability to work under pressure. Using this data, they compared exam results for those in the teaching profession to student test scores. Authors did not find significant relationships between teacher ability and student achievement. It should be noted that while this study had limited findings, the design of the study had many limitations as it used male sibling scores as a proxy for female teacher scores (as these exams are only given to males in Sweden).

In another international study of teacher cognitive ability on student achievement, Hanushek, Piopiunik, & Wiederhold (2014), estimated the relationship between teacher skills from the Programme for the International Assessment of Adult Competencies (PIAAC) and student achievement scores on the Programme for International Student Assessment (PISA) using data from 31 countries. While there are many different influences that could impact student achievement among several countries, the authors adjusted for these by applying multiple control variables to their OLS models, student-fixed effect models, as well as placebo effects to their analysis. Their findings show a positive significant (but minimal) relationship between teacher cognitive ability and student achievement as an increase of one standard deviation in teacher skills can increase student achievement in reading and math by approximately 0.1-0.15 standard deviations (at the $p<0.01$ level). Authors also noted that increasing teacher cognitive skills to that of the highest country in the sample (Finland) would have a similar positive effect, though would have the biggest effects for countries at the bottom. For instance, in Turkey and Chile, they estimate achievement scores in math to increase by 0.54 and 0.57 standard deviations.

2.1.8 University Selectivity

While the findings are mixed, there is limited evidence that a teacher's cognitive skills can have a positive (though small) influence on student achievement. As a result, researchers have often questioned whether the selectivity of the university where the teacher is trained could be another indicator of a highly effective teacher (arguing that selective colleges tend to admit applicants with higher cognitive abilities). Moreover, training from a more selective institution could be an indicator of a teacher who may possess higher teaching ability (thus the actual ability to teach, not only higher subject knowledge). Thus, university selectivity could be a proxy for a teacher's actual academic ability as well as her teaching quality (Kelly & Northrop, 2015)

While there can be positive and negative benefits from having more teachers from selective avenues, there is minimal, if any, evidence that the selectivity of the university where the teacher was trained has an impact on student learning. Researchers often argue that teachers with high cognitive skills are usually recruited from the most selective universities in the US (such as those who enter from TFA). However, Kane, Rockoff, & Staiger's (2008) study of TFA teachers from highly selective universities in New York City public schools did not find any notable differences in student achievement scores (when using institution GPA and SAT scores as a measure of institution selectivity). Furthermore, using data from North Carolina public schools, Clotfelter et al., (2007) also do not find any significant relationship between the selectivity of the institution a teacher attended and student achievement (for students in grades 3-5). However, when Clotfelter et al. (2010) adjust their model to include10 years of longitudinal data for students in North Carolina, they found small effects for student achievement in math (standard deviations between -0.007 and -0.010 at the $p<0.05$ level) for teachers that attended a non-selective institution (based on Barron's scale of university selectivity).

Authors Koedel & Betts (2007) also tested the relationship between competitiveness of the university where a teacher was trained and student achievement in math and reading. This study is notable in that the metric for university competitiveness was whether the university was a top 100 institution to receive research funding., instead of student GPA, SAT/ACT scores, admission rates, or other indicators of the university student population. Using student test scores in math and reading from the Stanford 9 standardized exam for students in grades 2-5 in the San Diego Public School System, authors found no relationship between the amount of research funding the institution where the teacher was trained and the achievement of his/her students. To further test the effects on university selectivity on student achievement, Clark Martorell, & Rockoff (2009) examined the selectivity of the university where principals received their bachelor's and master's degrees. Using student standardized exam scores for students in grades 3-8 in New York City public schools, authors found no significant relationship between the selectivity of the university where the principal was trained and standardized exam scores of students within the school where the principal presided. In Chingos & Peterson (2010) study of students and teachers in Florida, they also failed to find significant findings between the selectivity of an institution that a teacher attended and the student achievement of his/her students. They did find a significant negative relationship between student achievement on elementary math exams and teachers that attended competitive institutions (-0.017 standard deviations at the 0.01 level). However, they argue that though this was significant, there was only one grade with a significant relationship to the competitiveness of a teacher's university and thus could be inaccurate (caused by an unknown error in the data).

Goldhaber, et al., (2013) found inconsistent findings in their analysis of Washington State student and teacher data when they examined university selectivity and teacher effects. For instance, when they examined effects of a program's competitiveness (using the institution's average SAT score) authors found that competitive programs often produced less effective teachers and the opposite was also found to be true, as effective teachers were more often trained at non-competitive institutions. However, when they applied several measures of competitiveness (SAT scores, university admission rates, and average high school GPA) authors only found a minimal relationship between the selectivity of the institution a teacher attended and student math and reading achievement scores for students in grades 3-6 for years 2005-2006. For instance, when university selectivity was based on average high school GPA of admitted student and admission rates, significantly positive relationships were found between teachers at those institutions and student achievement. However, no other measure institutional competitiveness showed a significant relationship to student achievement.

Finally, Koedel, et al., (2012) ran a similar study using student achievement, teachers, and teacher preparation programs selectivity data from Missouri public schools. They did not report any significant findings between the selectivity of a teacher's training program and student achievement. They noted that teacher effects are more similar across different universities than average graduates from the institutions themselves. Such findings suggest there is limited variation between teachers when compared by institutions where they received their training.

2.1.9 Teacher Effects Beyond Student Achievement Scores

This review on teacher qualities and their effects (or lack thereof) on student achievement has primarily examined the literature that has used student value added scores from standardized exams. One reason for this is that value-added scores, while controversial and not without criticism, is often used in hiring, evaluations, salary decisions of teachers, and has begun to be used in other evaluations, such as teacher education (Rutledge, Harris, Thompson, & Ingle, 2014; Darling-Hammond, 2015). However, as valued added scores are only one measure of teacher effectiveness, it is important to briefly examine other measures to evaluate a teacher's effectiveness such as teacher ratings from principals as well as classroom observations.

In the literature on principal's ratings of teachers, one concern is that teachers ranked high from student achievement scores are not necessarily highly ranked by their principals (Harris, Ingle, & Rutledge, 2014; Rockoff, Staiger, Kane, & Taylor, 2012). One reason for this is that principals look for many factors in their evaluations of teachers that may not be directly related to student achievement but is still relevant (such as caring for student's well-being) (Harris, et al., 2014). However, in Harris, et al. (2014) study they noted that principal evaluations often identify those teachers with the highest student achievement gains. Furthermore, they note that while principal's value characteristics of teachers that are not necessarily apparent in value-added scores these are predictive of student value added scores. For instance, in survey and evaluation data, principals noted that teaching skills, motivation, team work, and enthusiasm were all important attributes of teachers. When these ratings were tied to value-added scores, they were statistically significant. Rockoff, et al., (2012) found similar relationships between value-added scores and principal evaluations, however, they also found that when principals received information about a teacher's value-added scores, principals were more likely to change their ratings. This suggests that principals do give weight to the scores and that objective evaluative methods can be combined with other evaluative methods for a richer review of performance.

Throughout the discussion of identifying an effective teacher, the conversation often points to the principal's observation of a teacher's classroom. These observations were historically the process by which teachers were evaluated and is often used for the majority of teachers where standardized test scores are not available. Classroom observations as an evaluative tool differs from some of the previous literature, in that it often is used to specifically identify *practices* of teaching quality, instead of the quality of individual 'teacher' attributes (Garrett & Steinberg, 2014). Researchers Kane, Taylor Tyler, & Wooten (2010) examined the role of the classroom observation in identifying effective teachers and their relationship to student achievement. In their study, they found that classroom practices measured in observations evaluations were predictive of student achievement. For instance, the higher the teachers score on classroom management, content focused teaching, and teaching skills, the higher the overall achievement of student in the classroom. Thus, they were able to identify common 'in classroom' practices that influenced student learning. They also found that teacher scores from observation were closely related to student achievement scores (as those with high evaluation scores likely had high student scores and vice versa).

In another examination of the relationship between classroom observations and student achievement, authors Garrett & Steinberg (2015) used the Framework for Teaching observation tools and examined whether results were correlated to student achievement scores. They found that for teachers with high classroom observation assessments, they also had higher than average student achievement scores. However, they also noted that this was contingent on the instrument used, as not all instruments measure the same teaching qualities and thus is difficult to state which measures influence student achievement. Both studies, Kane, et al., (2010) and Garrett & Steinberg (2014) argue that combined scores of value-added and classroom observations can work together for a more reliable prediction and estimation of teacher quality.

2.2 Teacher Training Programs

As the literature on teacher quality has indicated, there is no "silver bullet" to make a teacher highly effective at increasing a student's achievement. Some indicators of a more effective teacher has been a teacher's experience, her/his cognitive ability, and subject matter expertise, however these have shown marginal and inconsistent effects. As a result, researchers have explored whether the type of training a teacher as well as the teacher's training program itself had a relationship to student outcomes. Although the research to date has not shown an overall significant effect on student learning from the teacher's training program, there have been some unique findings. This section of the literature review will examine research on teacher training programs, paying close attention to studies that have tied student learning and achievement scores to a teacher's training. While there is an abundance of literature on this topic, this review will only look at those studies that have used state-wide datasets that link student achievement scores and the teacher programs where the teachers received their training. However, as teacher training in and of itself is made up of multiple components, this review will then shift to the more limited literature on components of teacher training such as quality of pedagogical understanding of teaching, foundations of teaching pedagogy, and student teaching experiences of teacher candidates.

One of the first studies to dive into the effects of teacher training and student achievement was, Boyd, Grossman, Loeb, & Wyckoff (2009). In this study, authors analyzed student and teacher record data in New York City using data on characteristics of programs, courses, and field experiences (provided through surveys of program completers). The researchers set out to determine if differences in types of teacher training (i.e., traditional versus alternative) resulted in different student achievement scores — and if so, which aspects of the training were correlated to higher student learning (Boyd et al., 2009). Findings showed that there was an institutional effect and that certain institutions produced more effective math educators and ELA educators. The differences were larger in math programs, however, with an increase of 0.18 standard deviations in math and a standard deviation of 0.10 in ELA.

Kukla-Acevedo, Streams, & Toma (2009) further examined the relationship between individual teacher preparation programs and student achievement by using teacher and student data for three districts in the Kentucky. Due to data limitations, researchers were only able to examine scores from students in 11th grade math courses for the years 2005-2006. Additionally, authors were only able to include three traditional university-based teacher training programs in Kentucky where most students received training, and then due to low sample sizes combined all other graduates from programs in Kentucky into one group, and a fifth group included those from out-of-state programs. Results in this study did not show any substantial predictive ability of individual teacher training programs on student achievement. Kukla et al. (2009) argued that the lack of significant findings related to teacher training could be a result of using all teachers instead of those with fewer than five years of experience and highlighted that there are several data limitations that hindered this study. Following this study, Chingos & Peterson (2010) examined of teacher administrative data linked to student achievement data in the state of Florida, and found significant differences for the University of Florida compared to all other programs (as teachers that received their bachelor's training there were significantly more effective). However, when the authors applied student and school fixed effects to their baseline model, all significant differences disappeared. There were two institutions that showed modest effects (one more effective at training elementary students and another less effective at training middle school students), however, authors cautioned that the sample sizes were very small. As a result, authors argued that overall there were no marked differences among teacher training programs in the state of Florida, for either bachelor's or master's degree programs.

In another study using data from Louisiana, researchers Gansle, Noell, & Burns (2012) examined variation among graduates of specific preparation programs by comparing student scores of recent graduates to those with three or more years of experience. Using math and English student scores from Louisiana state exams for grades 4-9 in the years 2006-2009, authors found that students of new teachers (with two or fewer years in the classroom) had lower scores than those with more experienced educators (-2.7 and -2.9 standard deviations at the 68 percent level due to low sample sizes) (Gansle et al., 2012). While this finding was true overall, there were differences among some programs, where new teachers showed a higher effect size across all content areas, and one had a higher effect size in all content areas. Additionally, three programs included in the sample showed significant effect sizes for new teachers in social studies and mathematics. Authors suggested that these results indicate there are differences in teacher effectiveness among programs (Gansle et al., 2012).

Koedel, et al., (2012) challenged findings from the Gansle et al. (2012) study and argued that authors used improper standard error clustering and therefore prior results were incorrect. Koedel et al. (2012) argued that because individual teachers have an effect on students, there was a clustering structure within the data (thus two students taught by the same teacher were not independent observations). To account for this, they clustered standard errors at the individual teacher level in this analysis. Using data from the Missouri Department of Elementary and Secondary Education on teacher certification, authors examined effects of traditional teacher training programs on student achievement scores in math and reading for fourth –sixth grade students during the years 2008-2011. Results showed there were no discernable differences in teacher preparation programs as demonstrated by student achievement scores. When authors examined the effects of clustering at the teacher level, findings confirmed that without clustering programs, large statistically significant differences existed across programs; however, when standard errors were clustered at the teacher level, variability existed within programs, yet no variability existed across programs Koedel et al., (2012).

Authors Mihaly, McCaffrey, Sass, & Lockwood (2013) also challenged previous findings and methodologies that tied student achievement to teacher training programs. Their research model tested whether findings were the same when controlling for school contextual factors and when school-level fixed effects were added to the model. Such model specifications were based on the assumption that new teachers were not randomly assigned to schools (but taught in areas near where they graduated) and that schools hired teachers from more than one teacher preparation program. Authors examined teacher data from those who taught in 4th or 5th grades between the years 2000-2004 and were graduates of a Florida elementary education preparation program. Findings showed that results were sensitive to model specifications as several programs moved between high and low rankings (based on student achievement data) when school fixed effects were added to the model. However, programs that were low ranked were more consistently ranked low, regardless of model specification. Authors noted the variance in program effects depended on the model and the number of years included in the data. They also argued that while the same programs consistently had a negative relationship to student achievement regardless of specification, too much variance exists for conclusive findings.

Goldhaber, et al., (2013) sought to further address the methodological issues noted in both Koedel et al. (2012) and Mihaly et al. (2012) studies. Goldhaber et al. (2013) tested the predictive ability of training programs when using student achievement scores and clustered standard errors at the teacher level and a school fixed-effects specification to their model. Authors also included a decay rate on teacher training effects to identify whether training impacts diminished after 10 years (Goldhaber, et al., 2013). Using teacher and student achievement data in math and reading for grades 3-8 for years 2005-2010, their baseline analysis found that most teachers who received certification from a Washington-based training program had small distinguishable effects, as less than 1 percent of teacher variation could be attributed to a teacher preparation program. However, similar to other studies, results varied based on the model specification used. For instance, in the decay model, the standard deviation of teacher effect was 0.20 in math and 0.16 in reading, providing approximately 5-12.5 percent of variation in teacher effect from training (Goldhaber et al., 2013). In adjusted models, they found that the difference between teachers from the least and most effective programs in math was 3.9 -13.4 percent of a deviation and 9.2-22 percent in reading. Furthermore, authors found that for both math and reading the estimated decay parameter was 0.05, suggesting that training program effects decayed as teachers gained more classroom experienced. They argued that evidence of decayed training effects suggests training does impact teachers quality (Goldhaber et al., 2013).

Researchers have used a similar analysis to examine data on teachers trained at institutions in North Carolina. Authors Henry, Patterson, et al., (2013) sought to compare student achievement scores of students whose teachers were trained at all public universities in North Carolina to those who were trained at other universities or routes in North Carolina. Using longitudinal data from statewide standardized test scores in North Carolina, authors found similar results to those found in Louisiana: there was significant differentiation within institutions and programs, but less between different institutions. For instance, one institution demonstrated a significantly positive effect in high school science and social studies (with an effect of 20.20 and 6.6, respectively) but was underperforming in middle school algebra (with an effect size of -13.5) (Henry, Patterson, et al., 2013). Of the 15 institutions in the study, three showed positive results in more than one tested subject area (yet one of those had negative results in another subject), seven had positive effects in one tested subject area, three had no significant results, and two showed only negative results in at least one tested subject.

In a similar study, authors Henry, Campbell, et al. (2013) examined effects of different types of teacher training programs on student achievement (and as such all institutions were grouped according to type and were not individually assessed). Of the 10 different teaching "portals" authors defined for this study, they used the in-state public undergraduate training portal as a reference category to all other entry portals. Authors found noticeable differences among entry types. For instance, teachers trained through out-of-state undergraduate programs had significant negative effects on student achievement in elementary, middle and high school grades (with an effect of -0.037 standard deviations for achievement in fifth grade science and -0.044 standard deviations in high school social studies achievement, both at the 0.01 level). However, similar to other findings, teachers from TFA programs had a significant positive effect on student achievement in elementary, middle and high school STEM with an increase of 0.058 standard deviations in elementary math, 0.188 in eighth grade science, and 0.223 standard deviations in high school science (all at the $p<0.001$ level). Overall, findings showed that graduates of traditional in-state programs outperformed graduates of other entry portals. Authors argue such results suggest there are significant differences between programs and types of training but cannot define attributes of specific programs that are responsible for the differences.

While much of this literature is focused on teacher training programs and their effects on the achievement of teacher's students, Goldhaber & Cowan (2014) examined teacher training and teacher attrition in Washington state. Using teacher data from public schools linked to data from the Department of Labor for the state of Washington state (for years 1989-2012), authors found significant variation in attrition among programs. They noted that while on average 15 percent of the teaching workforce in Washington leaves the profession each year, the differences in programs have a range of more than five percentage points, and the overall program effect was 1.5 percentage points in attrition at the 0.01 level, noting that there was a significant program effect (though the effect was more profound in small programs than at large public institutions). They further note, that much like the literature tying programs to student achievement, variation is more prominent within programs than between different institutions, suggesting the effects are more program based than the specific institution that houses the training program.

In addition to the literature testing the effects of teacher training on student achievement, two studies have examined the magnitude of findings in this literature. In 2016, authors vonHippel et al., used a data set of teacher data linked to student achievement scores for students in math and reading in grades 3-10 for the entire state of Texas (using 2011 data), to test effects of teacher training programs when different statistical models were applied. The variations in models they applied were: clustered standard errors, teacher random effects, school random and fixed effects, adjustments for multiple comparisons using the Bonferroni correction method, among other methods. Authors found that while there were significant differences between some programs (and an effect size of 0.03 standard deviations in math and 0.02 standard deviation in reading at the 0.05 level), there were very small effects and estimations were not reliable due to noise within the model specifications. Findings were not consistent between models, and noise accounted for between 50-90 percent of findings.

In a follow-up study, authors vonHippel & Bellows (2018) replicated datasets that were used in Texas, Washington State, Missouri, New York City, Florida, and Louisiana and analyzed the point estimates and standard errors that were used in each study. Authors argued differences among programs that were examined, differences were more pronounced in smaller programs than in large flagship programs. However, they argue the average program effect was an increase of 0.03 standard deviations (and the 0.05 level) which although significant, is minimal for actual improvements in student achievement. The authors did find that in some states there were significant and large differences between some programs, and approximately two-three percent of programs in all states were significantly different. Authors caution that while overall effects of training programs are minimal, and findings do not indicate why a program is different, there are some programs that have a significant effect on student achievement.

2.3 Elements of Teacher Training Programs

Much of the literature on teacher training has tried to identify if there are distinguishable attributes between programs, examining the effects of training on student achievement, and to a lesser extent, teacher attrition. However, recent attention has begun to merge what is known about teacher quality (such as experience) and teacher preparation to examine what attributes of a training program might (or might not) influence a teacher's effectiveness (and subsequently student achievement and teacher attrition). While this body of research is still developing, areas that have the most research are training programs in pedagogy (such as fundamentals of teaching), subject content pedagogy (beyond fundamentals of math, but instead the fundamentals of how to teach math), and student teaching.

2.3.1 Pedagogy and Teaching Foundations

One component of most traditional (and even alternative) teacher training programs is foundations of teaching and pedagogical theory. While program requirements vary between program types and institutions, training programs almost always require courses on teaching methods. However, literature on the effect these courses can have on student achievement scores are mixed, with limited evidence these are necessary (Allen, 2003). For instance, authors Voss, Kunter, & Baumert (2011) tested the importance of pedagogical knowledge by administering an assessment of a teacher's knowledge of teaching methods. Researchers then linked results from teacher assessments to student ratings of teachers. Researchers found a mild (but significant) correlation between a teacher's knowledge of teaching methods and student ratings of that teacher (r=.22). Authors argued that this was evidence that a teacher's grasp of teaching (regardless of the content) is necessary for student engagement. In a survey of middle school principal satisfaction with new teachers, researchers Howell, Cook, and Faulkner (2013) asked principals what areas new teachers needed the most assistance. Principals noted that while most new teachers came in with strong content knowledge, they were weaker in teaching methodology and pedagogical knowledge. Almost all principals felt confident their new teachers could teach their content area but did not think they were prepared to manage a classroom, especially a classroom of middle school students with changing needs.

Researchers Youngs & Qian (2013) found a nuanced understanding of the role of general pedagogy courses in a study of elementary teaching training programs in China. In this study, they found that the more diverse the general pedagogy courses were, the stronger the candidate's pedagogy content knowledge (with an effect of 0.778 standard deviations at the 0.05 level). Thus, having pedagogy courses that gave students more opportunities to learn teaching techniques and student assessment helped teachers strengthen their abilities to teach their content area. Authors also found a strong relationship between teacher preparedness and the level of their pedagogical content knowledge. A similar study of pre-service teachers in Germany examined their pedagogical content knowledge prior to and throughout their training programs. König (2013) administered a pre-service training test to primary level teachers and a post-service exam and found large gains in elements of teaching throughout their training. For instance, candidates had large initial gains in understanding and analyzing students and those gains continued through the end of their training ($d=1.33$ after the first two years and $d=0.37$ after the last two years). Similarly, for the metric 'generate knowledge' the author noted a gain of 0.59 for teachers in their first two years and a further gain of 1.29 for the last two years of the teacher's training. While this was the expected, authors argue that this finding is important because it confirms that pedagogical courses and student teaching should be paired as pre-service teachers need both to become effective.

2.3.2 Pedagogical Training in Subject Content

As prior research has suggested, a teacher's content knowledge can have a positive effect on student achievement. Such findings have led researchers to question whether teacher training programs should provide more content training to increase a teacher's effectiveness. However, this literature has also shown that content training alone is not enough, but instead subject matter training should be coupled with pedagogical methods to actually influence a teacher's effectiveness (for instance how to teach math to elementary students instead of just principles of teaching or more courses in mathematics). One early study sought to isolate effects on student achievement when teachers have subject-matter knowledge. Authors Hill, Rowan, & Ball (2005) examined teacher scores on a measure of mathematical knowledge for teaching tied to student achievement tests (for kindergarten through 5th grade) between the years of 2001-2004. The survey on teacher knowledge included questions about time spent teaching math, background in math, professional development in teaching math, as well as other educational activities. Authors found that a teacher's knowledge of teaching math had a strong predictive relationship to student achievement in math, with an increase of 0.05-0.23 standard deviations in student achievement. Authors also tested if this effect was due to a teacher's general knowledge of teaching or if it was due to knowledge of subject-specific teaching. Findings confirmed that only a small gain in student learning resulted from a teacher's knowledge of teaching methods, and that mathematics achievement had a positive relationship to a teacher's knowledge of teaching math content.

Harris & Sass (2007) continued this analysis and examined the relationship between required courses in subject matter teaching in teacher preparation to gains in student achievement in Florida for students in grades 3-10. Their findings showed that certain content courses offered in training programs had a positive effect. For instance, subject matter courses in mathematics (taught in the school of education with attention to pedagogical methods) had a strong positive relationship to student achievement in high school mathematics (effect size of 3.6037 at the 0.05 level). However, mathematics and statistics courses outside the school of education (for instance within the math department) had a negative effect, suggesting that subject content should be combined with subject matter pedagogy to increase teacher effectiveness. In a similar study, authors Henry, Campbell, et al. (2013) examined elements of teacher training that could influence the effectiveness of teachers and their training programs. To test program elements, researchers used data of teachers from a large public university tied to the elementary students (grades 3-5) of graduates from that program. By isolating one institution's effects, they explored attributes that could influence student achievement such as: content courses, pedagogy courses, subject specific pedagogy, as well as other academic metrics such as GPA and SAT/ACT scores. There were few significant findings in this study, but they did note a negative effect of one additional could on pedagogical methods of math (with an effect of -0.053 at the 0.05 level). This contrasts with the finding that the increase of one more content specific course had a positive effect on student learning in math (0.033 at the 0.05 level). Authors argue this reduction could signal that students took too many courses in one area at the cost of taking another that would've heightened the teacher's effectiveness.

Researchers Grobschedl, Mahler, Kleickmann & Harms (2014) examined various learning opportunities provided to teachers and questioned which opportunity (teacher training, professional development, on-the-job training) has the greatest impact on a teacher's pedagogical content knowledge. This study of secondary school teachers of Biology in Northern Germany administered an exam to test a teacher's understanding of subject content as well as the teacher's pedagogical knowledge of biology content. Results showed that when separated by training programs teachers had differing levels of content knowledge in biology (effect of 0.74 standard deviations at the 0.001 level), and those with additional professional development, have more pronounced content knowledge. Additionally, when examining the training type and its relationship to pedagogical content knowledge and content knowledge, no statistical differences were found between entry level teachers and those with more experience. Authors argue these findings suggest that content knowledge and pedagogical content knowledge is not gained by more experience but is instead gained through training programs. One limitation of this study is that the study does not offer insight into training programs or attributes of the programs to highlight possible reasons for these distinctions in teacher outcomes.

In a further study of teacher preparation and subject matter pedagogy training, researchers Ingersoll, Merrill, & May (2014) examined the effects of pedagogical courses within a teacher training program and their effects on teacher attrition using a nation data set on teachers in the US (the Schools and Staffing Survey 2003-2004 and the Teacher Follow-up Survey 2004-2005). Authors found that the type of training, degree type, certification and well as other characteristics (such as academic ability) had limited influence on a teacher's attrition. However, they noted that those most likely to leave were math and science teachers, who had more content specific courses in their subjects but had fewer pedagogy courses. Their findings also showed that those without strong pedagogical courses were three times more likely to leave the profession (at the 0.001 level).

Authors argued that their study suggests that content alone is not sufficient for teacher attrition but should be paired with pedagogical expertise.

Although most studies use student achievement scores to test the effects of a teacher's training in pedagogical content knowledge, one study also examined a student's interest in the subject (Keller, Neuman, & Fischer, 2017). Authors examined effects of a teacher's pedagogical content knowledge using a rich dataset that combined teacher pedagogical content knowledge in physics, videotapes of instruction, student achievement data, and teacher and student surveys on motivation and interest in the subject. Findings showed that student achievement and interest were correlated with r=0.25 at the pre-test and r=0.33 for the post-test at the 0.001 level. Furthermore, results also showed that a teacher's pedagogical content knowledge was predictive of student achievement (with an effect of 0.44 of a standard deviation), and teacher motivation was predictive of student interest (with an effect of 0.53). However, a teacher's pedagogical content knowledge in physics was not related to teacher motivation, and a teacher's pedagogical content knowledge did not have a relationship to student interest in physics.

In a similar study, author Preston (2017) examined the relationship between components of teacher training programs (such as content courses, pedagogy courses, and seminar) and student achievement for middle grade students in reading and math. Preston (2017) found few meaningful results, except that subject matter course work was negatively related to student achievement. The author noted that an additional one-hour credit hour course in math content had a decrease of 0.0043 standard deviations (at the 0.065 level) in student achievement and an additional three-hour course in math content had a decrease of 0.0129 standard deviations. The author noted that math teachers take additional courses in mathematical content, suggesting that teachers could be missing other important courses that influence a teacher's effectiveness.

2.3.3 Student Teaching

While much of this review has examined individual courses within training programs, one essential component of teacher training programs is student teaching. Extensive literature has examined elements of student teaching that could influence a teacher's effectiveness such as: the cooperating teacher, placement of the student teacher, the length of the experience, feedback and observations of the student teacher, and the impact of student capstone courses. While this is an area with rich data, the training program itself often has limited influence on elements of this experience. As such, this section will primarily examine the literature on capstone courses, the length of the training, and the overall effects noted from the student teaching experience (noting that several of these elements are beyond the control of the training programs themselves).

In 2009, Boyd et al., examined attributes of teacher training programs and found that observations of student teaching and capstone courses had the largest impact on first-year teachers (observations had a coefficient of 0.03 in math and 0.01 in ELA, and capstones had a coefficient of 0.04 for both math and ELA teachers). These findings suggest that there is a positive relationship to student achievement when the teaching candidate's student teaching experiences are paired with feedback on his/her instruction (Boyd et al., 2009). Building off of the literature on a teacher's pedagogical content knowledge, authors Friedrichsen et al. (2008) examined the role of student teaching in the development of a teacher's ability to teach biology. In this study, some participants had student teaching experience while others only had training in teaching methods in their courses. Authors found positive significant results when teacher content knowledge and teaching ability were paired with in-classroom training and curriculum instruction on how to teach.

Authors Ronfeldt & Reininger (2012) tested the hypothesis that more experience teaching influences teacher confidence and ultimately a teacher's retention. In this study, authors examined if the length of the student teaching experience had the same effect as years of being the teacher of record in a classroom. Through a pre-and post-test survey of student teachers, authors examined how prepared student teachers felt to teach students and to modify lessons if needed. They found that length of student teaching had no effect on a prospective teacher's confidence or preparedness, but instead satisfaction with the cooperating teacher was a bigger predictor of a student teacher's preparedness (an effect of 0.139 at the 0.001 level). In another study, Ronfeldt (2012) tested whether easy to staff or difficult to staff field placement schools had an effect on a teacher's retention and student achievement. The author was able to tie student achievement gains and teacher retention to field placement schools in New York City public schools. It was found that teachers trained in 'easy to staff' (those with higher teacher retention rates) field placement schools for their in-service training had higher retention rates (effect of 0.6904 of a standard deviation at the 0.001 level) and higher student achievement gains (effect of 0.0309 increase in a standard deviation at the 0.05 level) than those trained in hard to staff schools. The author argued, while the teacher training program is often not responsible for the placement, helping to place candidates in easy to staff field placements could positively influence a candidate's retention.

In a study of Chinese elementary teaching candidates, authors Young & Qian (2013) examined survey data of candidates to better understand their development of pedagogical content knowledge of mathematics. They examined candidate progress in pedagogical courses, math courses, and in student teaching. They found that the number of courses taken in mathematics and the length of student teaching hours had no effect on the teacher's knowledge of teaching mathematics. However, they did find that student teachers with longer periods of full responsibility of their training classroom (compared to just assisting or observing) had a positive impact on the teacher's development in teaching math, specifically in teaching numbers and operations (an increase of 0.53 standard deviations at the 0.05 level). Authors Ronfeldt, Reininger, & Kwok (2013) sought to further disentangle student teaching from university training by analyzing pre- and post-student teaching surveys. In this analysis they questioned the effects that student teaching and general teacher characteristics had on a teacher's perceived instructional quality. It was found that career plans (and subsequent longevity in the profession) are more related to general teacher characteristics and demographic elements (such as training location), than the student teaching experience. However, they also found that student teaching experiences influenced whether candidates saw themselves as effective or not. Specifically, authors found a positive relationship between the cooperating teacher and the student teacher's confidence in his/her preparation and effectiveness (with an increase of 0.17 standard deviations at the 0.001 level for preparedness and an increase of 0.199 standard deviations at the 0.001 level for efficacy). There was also a positive relationship between the amount of autonomy the student teacher had and his/her perceived preparedness (with an increase of 0.16-0.18 in standard deviations at the 0.001 level for preparedness).

When authors Ingersoll, Merrill, & May (2014) examined the effects of student teaching on first year teacher retention, they found that first year teachers without in-service training were three times more likely to leave the profession (at the 0.001 level). However, those with at least 12 weeks of in-service training were two times more likely to stay in the profession (and the 0.05 level). In a following study, researcher Ronfeldt (2015) examined the school where a teacher candidate received her/his training to test if that experience had an effect on student achievement. The author used candidate surveys, data on student teacher placements, achievement scores of students later taught by those candidates, and in-service placement school characteristics (such as teacher turnover, suspensions, achievement scores, and other descriptive and demographic data).

It was found that in-service experiences that are collaborative led to student achievement gains of 0.094 standard deviations at the $p<0.05$ level, as did those prepared in field placements with less teacher turnover (with a coefficient of 0.07 at the $p<0.10$ level). Furthermore, they found that reading teachers with in-service experiences at the same grade level as their first-year placement had higher student achievement gains, than if their training was at a different level (with gains of 0.17 standard deviations at the 0.01 level). Similar to other studies the length of student teaching was not associated with student achievement.

Finally, author Preston (2017) examined the effects of student teacher in-service experiences and effects on student achievement gains. She found that a 10 hour increase of required in-service training in mathematics led to a decrease of 0.003 standard deviations in student achievement gains (at the 0.001 level), when the student teaching was early in the teacher's training. Furthermore, additional hours of teaching (and not just assisting the teacher of record) led to a decrease of 0.01 at the 0.01 level. It was found that a capstone/seminar course at the end of the training experience lead to an increase in student achievement in math (0.05 at the $p<0.1$ level). However, this finding was not confirmed when other models were tested, thus the author was hesitant to suggest seminar courses were predictive of student achievement.

2.4 Qualitative Review of Teacher Training

This literature review has primarily examined literature on quantitative assessments of teacher training programs, using student achievement scores and teacher attrition as a measurement of a program's effectiveness. However, many authors have argued that such assessments do not capture the full effects of teacher training as they are only able to measure increases in student achievement scores and not whether the training program itself (or components within the programs) was the determining factor. Furthermore, they argue the research shows there are more differences within programs at the same institution, and thus quantitative measures do not capture features that accounts for these differences. As a result, many researchers have called for a mix of evaluative methods that combine quantitative student achievement data with teacher survey data and other qualitative metrics (Darling-Hammond, 2016). For instance, Chung (2008) performed a study of pre-service teachers through surveys, interviews, case studies, and of candidate responses to teacher performance assessments. The author found that candidates with more opportunities to assess student learning (and differentiate lessons for student needs) in their pre-service experiences were more likely to continue the practice once they were the teacher of record. Through case studies, the author identified specific training such as independently crafting lessons and adjusting lessons for individual student needs that filled gaps present in subject's traditional training. Furthermore, case study and interviews provided detailed information about the lessons and experiences within the student teaching experience that worked, as well as those that did not help in developing the candidate's practice.

In another analysis, Darling Hammond, Newton, & Wei (2010) reviewed research on teacher candidate surveys and interview data from the Performance Assessment for California Teachers (PACT). This review highlighted the benefits of qualitative mechanisms to evaluate teacher training programs. For instance, they noted that candidates that felt the most prepared to teach were also more likely to respond that they adjusted teaching based on student progress and needs, made data-informed teaching decisions, and sought out new opportunities for learning. They were also able to identify more granular instances where courses directly affected teaching practice. In another example, authors interviewed candidates about the effects of a course redesign that intentionally linked their weekly student teaching experiences to class readings. Teacher candidates noted this redesign was more effective and helped them develop their teaching practice. Finally, as noted previously, knowledge of teaching can have a positive impact on student achievement. Darling-Hammond et al. (2010) examined research that used pre-and post-test on candidate knowledge of teaching and combined this with candidate interviews. The data helped researchers identify only areas of weakness from the training and also dive into areas where candidates needed more training (such as additional readings or practical applications).

More recent literature on qualitative assessments of teacher training and components within teacher training has focused on the edTPA. EdTPA is a teacher performance assessment administered to teacher candidates to assess their teaching competencies. The goals of this assessment have been to identify a teacher's teaching ability and their understanding of teaching methods. Authors Ledwell & Oyler (2016) sought to compare candidate data from the edTPA (from training programs in New York state) with interview data of teacher candidates as well as teacher educators. In their study they found that while programs did not always know how to make changes when candidates received low scores on the edTPA, programs did modify curriculum as well as instructional methods. However, interviews highlighted that much of the changes were superficial and at times compromised reflective opportunities that candidates had previously found useful. Authors cautioned that while standardized assessments such as the edTPA can be useful to point to areas of weakness (or to highlight areas of strength), without qualitative assessments training programs may not know areas to improve and more importantly, aspects that candidates find valuable and necessary when developing their craft.

2.5 Conclusion

As data quality has improved over the last few decades, researchers have been able to study the various aspects of teachers that may make them more (or less) effective. However, even with improved datasets and more nuanced research, research has yet to find the 'silver bullet' of teacher quality. For instance, some research argues that teachers with strong pedagogical content knowledge and more years of experience are likely more effective. Furthermore, some research has found that certain training programs or certification programs (like TFA) have higher quality teachers. However, there is limited consensus on this research and most findings are not generalizable to all teachers, but instead are relevant to the particular context within the specific study, and at times are dependent on the model specifications used in the analysis (such is the case with the teacher preparation literature presented). That said, the literature does agree that teachers matter and there is some agreement that training matters. However, individual aspects and contexts are dependent of many variables still to be realized.

3.0 Methods

Using a quantitative approach that draws on qualitative methods, this study aimed to identify what, if any, elements within an elementary teacher preparation program have a relationship between elementary student achievement gains in reading and math as well as a teacher's attrition. This study used findings from interviews, program document reviews, and surveys of training programs in North Carolina to identify what elements of program quality, in addition to those publicly available, are present in teacher preparation programs. I then analyzed longitudinal data on elementary student achievement gains as well as teacher attrition in the 2014-2018 school years from the North Carolina Education Resource Data Center (NCERDC). I then tested program elements and their relationship to elementary student achievement gain scores and teacher attrition using a covariate adjusted multilevel linear model as well as a school fixed effects model. The research questions that guided this study were:

1) What program elements, beyond those available in the public domain, are present in teacher preparation programs and are indicators of program quality for elementary teaching programs in North Carolina?
2) What is the relationship between elements of an elementary teacher preparation program and elementary student achievement gains in math and reading in North Carolina?
3) What is the relationship between elements of a teacher preparation program and elementary teacher attrition in the first 8 years of a teacher's career in North Carolina?

This chapter will first walk through the research design process that was used in this study. I then provide a detailed description of the data used as well as the analytic sample of students, teachers, and programs examined in the study. Finally, I describe the analytical approaches used to answer each research question. The last section of this chapter provides details on data collection (including interviews, document review, and program surveys) as well as the statistical methods used to assess relationships between program elements, student achievement and teacher attrition.

3.1 Research Design

3.1.1 Methods to Identify Teacher Training Program Elements of Quality

Prior literature on elements within teacher training programs and their relationships to student achievement and teacher attrition have suggested that while differences between programs may exist, elements that are not publicly available may have a stronger relationship to student achievement and teacher attrition than those identified through public sources (Henry, Campbell et al., 2013; Preston, 2014). As such, this study used multiple approaches through interviews, document reviews, and surveys of training programs to gain a broader and more complete picture of elements of teacher training program quality (Mertens, 2005).

I first interviewed select training programs to understand local and state context of the elements within programs and the programs themselves. Although prior literature has explored broad program elements, program culture and state context are necessary to understand the programs themselves. Interview findings and literature on training programs provided the foundation for me to develop a survey of program training techniques and methods. Through program surveys, I collected information from identified programs and then quantified those elements for use in the statistical analysis. Programs were selected to be in the survey sample only if the program showed a significant statistical relationship to student achievement gains in an initial analysis. Survey responses were coded into variables to be used in the quantitative approaches listed below (Nardi, 2018). Note that, unfortunately, the sampling strategy (designed to facilitate a qualitative understanding of effective programs) reduced the total variance in program elements identified by the survey, and thereby reduced statistical power in the quantitative analysis of program element effects. Thus, my analysis provides somewhat conservative statistical tests of program element effects. Finally, data from program surveys was also paired with data from public documents on training elements such as program curriculums and Title II reports on teacher training from the US Department of Education. Although the focus of the study was to understand elements within the programs (and thus those beyond what is publicly available), these public elements may have also demonstrated a relationship to student achievement and/or teacher attrition.

3.1.2 Quantitative Methods

Once elements of program quality were collected, I employed quantitative methods to test the relationship between those elements, student gains, and teacher attrition. However, several methodological concerns were taken into consideration before I proceeded. For instance, the purpose of causal research at its most basic level, is to examine if the occurrence of A (or presence of A) *caused* B outcome. In this study, causal research measures how student (and teacher) outcomes may differ based on the existence (or absence) of a training element (Murnane & Willett, 2011). As Preston (2014) argued, the ideal experimental research design to investigate the effects of elements within a teacher's training program on student achievement would be to measure the teacher's value-added score[6] without having the training characteristics, and at the same time, his/her value-added score with the training. If this were possible, one could then measure effects of the training on student achievement by:

$$\boldsymbol{TE_i = Y_i(1) - Y_i(0)}$$

Equation 1

Where the training effects of teacher (i) TE_i would be determined by finding the difference between the value-add score of the teacher with the training $Y_i(1)$ and his/her value-add score without training $Y_i(0)$ (Murnane & Willett, 2011). This would ensure that all variables could rule out other variables related to student achievement, and thus only the training variable is responsible for the differences (Kelly, 2016; Mertens, 2005; Murnane & Willett, 2011). However, this design is not possible as a teacher's value-added score cannot be measured at the same time with and without training.

Since true experimental design cannot be applied in this study, a quasi-experimental research design method had to be performed. In this design method, the key independent variable (elements of teacher training programs) was measured using methods that account for that variable, the dependent variable (student achievement), and variables related to both (the confounding variables) (Kelly, 2016). For this study, one place to start is to apply a standard linear value-added model to estimate individual teacher (or training) contributions to a student's achievement. In this model, the use of a student's gain in achievement as a key variable serves as a statistic of previous

---

[6] Unless otherwise noted, value-added scores are defined as a measure of each student's human capital accumulation at point a and then the estimated changes to this human capital resulting from inputs obtained at time period b (Koedel, et al., 2015).

inputs to the student's ability and can therefore be used to estimate achievement when future inputs are added to the model (such as components of teacher training programs). This model can also accommodate covariates to strengthen the predictive ability of the model and identifies competing explanations of relationships between the key variable (Clotfelter, et al., 2010; Kelly, 2016; Koedel, Mihaly, & Rockoff, 2015; Koedel, et al., 2012; Preston, 2014). The standard linear model to test program effects on student achievement is:

$$\boldsymbol{Y_{icjst} = \beta_0 + \beta_1 Y_{icjst-1} + \beta_2 X_{icjst} + \beta_3 C_{icjst} + \beta_4 J_{icjst} + \beta_5 S_{icjst} + \theta P_{icjst} + \varepsilon_{icjst}}$$

Equation 2

In this model, Y_{icjst} is the student achievement score for student *i* in classroom *c*, for teacher *j*, in school *s*, at time *t*.

$Y_{icjs(t-1)}$ is student *i*'s achievement gain,

X_{icjst} is a vector of demographic and socioeconomic data for student *i*

C_{icjst} is a vector of classroom characteristics for classroom *c* in school *s*,

J_{icjst} is a vector of teacher characteristics

S_{icjst} is a vector of characteristics of the school attended by student *i*

P_{icjst} is a vector of training elements within teacher *j*'s teacher preparation program (and are the key variables of interest)

ε_{icjst} represents a random error term.

In equation two, the model examines a student's achievement as the summation of a cumulative set of inputs based on a student's prior knowledge and includes θ as a coefficient of P_{icjst} (to denote elements of training programs as the key variable of interest). Instead of the randomized experiment model in equation one, this model statistically measures the relationship between program training elements and student achievement, with the inclusion of covariates.

One limitation of this model however, is that simply adding a program element variable along with other variables to the model, does not necessarily isolate the effects of the program itself. For instance, if the relationship between components within a program and a student's achievement is due to the presence of a mediating or confounding variable, the model must also examine this relationship to understand the effects of training elements on student achievement (Kelly, 2016). Another consideration of covariate adjustment is that covariates can only predict variation that is not already predicted by other covariates. Thus, if the relationship between program elements and covariates are correlated (which in this study they likely are), methods must be applied to measure and account for this correlation, otherwise estimates may become less predictive (Murnane & Willette, 2011). Furthermore, simply adding these covariates does not account for bias within the variables (from non-random sorting as well as intentional placement and selection of students, schools, and training programs) nor does it consider the nesting structure of the data (which can result in inaccurate estimates) (Allison, 2005; Bryk & Raudenbush, 1992; Henry, Bastian, et al., 2014; Kelly, 2016; Koedel, et al., 2015; Luke, 2004). To adjust for these limitations in equation two, authors suggest two statistical methods: a covariate adjusted multilevel model and a school fixed effects model (Allison, 2005; Henry, Bastian, et al. 2014; Murnane & Willett, 2011).

Considering the above concerns with the linear value-added model in equation two, one recommended approach is to apply a two-level hierarchical linear model[7] that adjusts for student, teacher, and program covariates. Multilevel models are applied to data collected at more than one level (such as student, teacher/classroom, school or program levels) to identify relationships that

[7] Hierarchical linear models are referred to as either hierarchical linear models or multilevel models.

exist between the predictor (classrooms) and the outcome (student achievement) variable by accounting for variance amongst variables at multiple levels (Luke, 2004; Woltman, Feldstain, MacaKay, &Rocchi, 2012). In a multilevel model, each 'level' is represented with its own sub-equation. These sub-equations estimate relationships between variables at each level and specify how variables at one level influence relationships at another level (Bryk & Raudenbush, 1992).

A key strength of any multilevel model is that by recognizing the nesting structure of data (here students within teacher-classrooms), observations are not independent of each other. Woltman et al., (2012) explains that the two models examine effects within individual units and test cross-level effects. For instance, one examines the relationship of student variables only, and the second level model (teacher-classrooms) estimates the relationship of students between both students and teacher-classrooms. This is done by estimating each level 1 unit (student) by its level 2 cluster (teacher-classroom), and then the slope of each teacher-classroom is also estimated and analyzed. Residuals are then partitioned into a student error term and a classroom error term. All students in the same teacher's classroom share the same value of the classroom residual and this correlates with their combined residual. This approach accounts for the intercorrelated residuals as they are not independent from one another but are assumed to be distributed independently in the population. Furthermore, because the outcome variation is separated into student and classroom components, between-school and within school variance is also modeled. However, not applying a multilevel approach would risk incorrectly partitioning variance increases (that is, the ability to correctly identify the correlation among student-level variables with-in and between classrooms decreases) (Bryk & Raudenbush, 1992; Murnane & Willett, 2011; Woltman et al., 2012). A final consideration in a multilevel model is to adjust for clustering within standard errors of students in the same classroom (von Hippel, et al., 2016). By not accounting for within group correlations (such as that of students and teachers) inaccurate standard errors, narrow confidence intervals, large t-statistics, and low p-values can occur (Cameron & Miller, 2015; Goldhaber et al., 2013; Henry, Bastian et al., 2014; Koedel et al., 2012; Mihaly et al., 2013, von Hippel et al., 2016).

Although a multilevel approach considers the nesting structure of data (and thus the non-independence of observations from their classrooms), bias is still present in the model (Henry et al., 2014). Authors Clotfelter, Ladd, & Vigdor (2010) note, that due to personal preference, location, and school policies (such as tracking) students are not randomly assigned to their classrooms and teachers are not randomly assigned to those classrooms or schools. This lack of random assignment can bias the results and overestimate the effects of training methods on student achievement (Koedel, et al., 2015). Authors Koedel, et al., (2015) argue that due to these limitations, researchers must employ multiple methods to address these biases such as reducing non-random sorting (where possible) and using multiple data points on students and their outcomes (such as longitudinal data). However, they also note that because these measures are difficult to implement, varying models such as those of a fixed effects model can be examined.

Through a fixed effects model, the researcher statistically controls for characteristics of higher-order units (e.g. schools or districts) that are unobservable (such as elements of a school's academic climate) and could be related to the dependent variable (student achievement) (Allison, 2009; Clotfelter, et al., 2010; Koedel, et al., 2015). Unlike a multilevel model, fixed effects models only examine within-group variation; all between-group variance associated with the group-ids is removed (Allison, 2005). Fixed effects models are useful when between-group variation is confounded with unobserved characteristics of individuals, such as when student sorting occurs between schools that cannot be fully-accounted for by measuring student-level variables, either due to a lack of available measures or measurement error (which is always present to some extent).

While fixed effects models aim to account for biases present in statistical models, one question is what between-unit variance should be fixed out or removed. For instance, in this study, bias is present between schools as teachers are not randomly sorted to schools, and within schools and between classrooms as teachers and students are non-randomly matched. Teachers are assigned to their schools through many different methods, few of which are random (Clotfelter, et al., 2010; Koedel, et al., 2015). Teacher assignments can be a result of school needs and teacher abilities (such as math expertise), or the school's location in proximity to the teacher's regional preference (Goldhaber, et al., 2013). Indeed, location itself accounts for much of the non-random assignment of teachers to schools, as many schools employ teachers who were trained at the same program. As such, the application of school fixed effects can limit the bias that is present from systematic differences across training programs in the k-12 schools where their graduates are placed (such as quality, socioeconomic status of students within the school, or underrepresented student populations), by allowing for within school comparisons only (Koedel, et al., 2012). In other words, a school fixed effects model compares the variation in student achievement both within schools across different teachers, and potentially over time as well in data with multiple years for each school (over time schools have different students and teachers, and teachers would have been trained through different training programs).

One concern with a school fixed effects model is that as Goldhaber et al. (2013) note, schools often employ teachers that are similarly effective, and as such their effectiveness may be the reason they are placed into a school. For instance, schools may employ teachers who are the least effective but graduated from the best training program, or the most effective teacher from the least effective program. As a result, they argue that within-school comparisons would show limited differences between program elements and student achievement and may introduce comparison group issues when interpreting the results (Allison, 2009; Goldhaber, et al., 2013; Koedel, et al., 2012; Mihaly, et al.; 2013). However, while there are in theory limitations to using school fixed effects in a study of teacher training, when researchers have applied school fixed effects to address the known bias in school placement, results showed minimal changes (Goldhaber, et al., 2013; Koedel, et al., 2012; Mihaly, et al., 2013).

Previous studies have also added a fixed effect at the university level, noting that while teacher candidates' sort into programs based on personal preferences (and thus intentionally select the program), they also sort into the university that houses the program. As such, those in the program are subject to the same university admission requirements, have the same university resources available to them, and may be similarly interested in university specifications (such as size or location). A university fixed effects model would compare the variation in student achievement over time taught by graduates of the same university as program elements changed over time (or perhaps compare programs within a university, but only if the outcome measure could be fully equated). However, there are major limitations to a university fixed effects specification in any study and certainly this study. The first is that there are often are no alternative within-university programs for teachers who will have the same teaching assignments. As this study focused on elementary teacher training, there is extremely limited within-university variation in the programs of the teachers in the dataset.[8] Overwhelmingly, for those that have their major listed, approximately 80 percent received a degree in elementary education from their university.

[8] There are programs where the elementary training program has a specialized focus such as math elementary education, however, distinguishing that in the dataset was not possible as training programs and degrees were not identified with that level of detail.

Second, individual universities don’t engage independently in major program revisions very often. Indeed, findings from the survey in this study showed that program requirements in recent years changed primarily as a result of a state-led intervention (which applied to all universities), and even then, actual changes were rare. As such, a university fixed effects model would not provide meaningful variation in program features in this study.

3.2 Sample Criteria

Data for this study was provided from the NCERDC, and as such, the population for this study was limited exclusively to teachers and students, in public schools in the state of North Carolina. Additionally, teachers (and students taught by those teachers) who were trained outside of the state of North Carolina were excluded from this study.

For the analysis on the relationship between student gain scores and elements of teacher training programs, only elementary students (grades 3-5) with gain scores[9] from state standardized tests (in math or reading) who were taught in public schools in North Carolina between 2014-2018 were included. Additionally, for this analysis, only teachers with student gain scores in math and reading were included. When identifying a teacher’s training program, the data file did not indicate the teacher’s training program, only the teacher’s degree, university they attended for their degree, and type of license. As such, I only included teachers whose license indicated they were trained at an institution of higher education and were licensed through their BA degree. All teachers with a

[9] Students had to have achievement scores in the same subject for 2 years to be able to calculate a gain score (Score Y2-Score Y1), thus students with only one achievement score were excluded.

license awarded due to training at a university were tied to their bachelor's degree university. While this did not substantially alter the sample size, it removed teachers who were trained through Teach For America (TFA) or alternative routes. Additionally, as teachers in charter schools in North Carolina are not required to have a license to teach, the majority of teachers in this study were those in traditional public schools. There were teachers in the file that had 1 or 2 years of experience but who graduated with a bachelor's degree prior to 2014, indicating a career-stoppage had occurred. After reviewing the data, only those that graduated with a bachelor's degree in 2011 or later were included (as lengthier career-stoppages among teachers graduating before 2011 but were still early career teachers during the study period, were far less common).

For the analysis on teacher attrition, the sample did not include student gain scores.[10] The criteria for teachers to be included in this sample was: a teacher who taught any elementary grade (1st-5th grade) between the years 2014-2018, were trained in North Carolina (with the same license criteria used previously to link teachers to a training program), had a record in the teacher salary file, and entered teaching in 2009 or later. Additionally, the salary file from which the sample was built only had data on those employed in public schools and did not include those who taught in charter schools, and therefore teachers from those schools were excluded.

Finally, for the purposes of this study, teacher training programs are defined as a bachelor's degree program at an institution of higher education that is approved to train teachers by the North Carolina Department of Public Instruction (NCDPI, 2018a). There were substantially smaller

---

[10] The decision to not include student gain scores in this portion of the study was because tying teachers to students with gain scores greatly reduced the teacher sample. As such, it was necessary to have a larger sample to track teacher patterns. Additionally, the focus of this study was on whether teacher attrition had a relationship to their training program and not that of student gains.

programs in the data and those were therefore removed as their small size would make it difficult to make inferences about program elements and effects. For instance, one small private program that typically produces less than 20 teachers a year was removed as it only had 5 teachers in the datafile (before any filters were applied). Comparatively, one program that was included was one of the largest public programs in the state that graduates approximately 10 percent of all teachers in the state and had over 1,000 teachers in the datafile. Given these differences in size, programs were initially eliminated if they had less than 100 total teacher records for the gain score analysis and less than 50 teacher records for the attrition analysis. These cut-offs were determined based on natural declines in the data of the programs and by what would ultimately yield a feasible sample size after all sample filters were applied.

3.3 Data

This study combined student and teacher level data with program data to answer the study's research questions. Data on program elements include those publicly available as well as data collected from surveys of training programs throughout North Carolina. Secondly, data from the NCERDC provided records on elementary student achievement in math and reading as well as the record of the teachers of those students. Details of each dataset are outlined below.

3.3.1 Teacher Training Program Elements of Quality

As noted previously, there is limited publicly available data on teacher training components, and as a result, I combined available data from the US Department of Education as well as program curriculums (from program websites) with a survey designed specifically for teacher training programs in North Carolina. Although this public data was not nuanced enough to fully evaluate what underlying elements of quality may affect student gain scores and teacher attrition, these sources did provide broad information on program requirements and were thus incorporated into the analysis.

3.3.2 US Department of Education Title II Reports

Under the Higher Education Act of 2008, states are required to provide the US Department of Education with data on teacher training programs within their respective state (US Department of Education, 2016c). This data primarily includes structural information on program requirements such as names of approved programs, licensure passage rates, among other programmatic information (for the years 2009-present). Much of the information in these reports is descriptive and does not identify programmatic elements that vary across universities. However, the Title II reports do include a few variables that differ across programs and can be tested in a quantitative model. The elements from these reports used in this study were:

- Supervising FTE Faculty and Adjunct Faculty:
    - Number of FTE faculty that supervise candidate teaching
    - Number of adjunct faculty that supervise candidate teaching
    - Number of students in student teaching (in current year)
- Student Teaching
    - Number of required program hours prior to student teaching
    - Number of required student teaching hours

3.3.3 Program Course Curriculums

In addition to data from Title II reports, I reviewed course catalogs of each elementary teacher program in the study sample. This provided further information on the training that program graduates received, such as number of math specific courses compared to number of foundation courses they took. In this review of program curriculums, I identified 8 course categories: subject-only courses, student assessment courses, content teaching courses, method-only courses, instructional technology courses, learner and classroom diversity courses, foundation courses, and early student teaching opportunities. I used the following rules to code each identified course:

- Math or English Subject-only courses:
  - Courses included general education[11] math, writing and/or English.
  - Additionally, courses in the school of education such as principles of elementary math and children's literature courses.

- Student Assessment Courses:
  - Courses that taught how to assess students either as defined in course descriptions or as labeled in the course title.

[11] Because programs included in this study were at the bachelor's level, general education courses were included for only math and English, all other general education courses were excluded.

    - In occasions where the course was has multiple categories within the description such as “classroom management and assessment” the course was counted as an assessment course.

    - In instances where the focus of the course was assessment of a specific subject, such as “assessment of reading students” that was counted as a content method course.

- Courses on Teaching Subject Content[12]:
    - Courses were counted in this category if it was distinctly noted either through the description or title that the course was focused on methods of teaching math or reading.

- Teaching Method Courses:
    - Courses included general teaching methods such as instructional methods, classroom management, or other teaching methods.

    - Courses focused on subject -specific methods or teaching special populations were not counted in this category.

- Instructional Technology Courses:

---

[12] Although programs had courses on teaching science and social studies, these were not included in this analysis as it was outside the scope of tested subjects.

    - These courses were counted when the focus was how to incorporate technology into their teaching practice.

    - In cases where the course had multiple categories within the description such as "instructional technology and assessment" the course was coded as instructional technology and not assessment.

- Special Learner Populations and Classroom Diversity Courses:
    - Courses were included if the focus was teaching special education, exceptional students, inclusion education, English language learners, or other student populations.

    - Courses with a focus on teaching in a culturally diverse environment were also added in this category.

- Education Foundation Courses:
    - Courses in this category include intro to education, courses on child psychology, education theory, or foundations of teaching.
- Early Student Teaching Opportunity:
    - This was noted if the program had an opportunity for the student to observe, co-teach or attend an elementary classroom in the first semester of training (as noted in the course description).

- Average Undergraduate SAT Score:
    - Although not specific to the program, this was included to test for institution competitiveness where the program was housed.

In some cases, program requirements changed during the study period. To account for this, I reviewed the last 3 years of curriculums and when there were changes, I took an average those changes. However, in most instances, courses had different titles or course numbers, while the courses themselves remained the same (for instance in one year there was a course on instructional technology and the next year it was titled instructional technology and assessment).

3.3.4 Data from Program Surveys

Although the program data from curriculum reviews and Title II reports provided a broad understanding of the training teachers from each program received, I wanted to identify training methods that occur within courses and the program itself to test whether those elements had an effect on student and teacher outcomes. For instance, while curriculum reviews highlight the number of math courses, it does not say whether candidates record their teaching, nor the process by which a candidate's teaching is assessed. To better understand techniques used within programs, I surveyed programs in North Carolina about their training processes (see Appendix A for survey instrument). From survey findings, I identified six teacher training categories:

- Faculty Teaching

- Questions included: percent of adjunct faculty, courses taught by faculty, incentives for innovative faculty teaching methods.

- Teaching Experiences Prior to Student Teaching
    - Questions included: in-classroom experience in the intro to education courses, number of courses that had active teaching experience (outside of student teaching), and whether the program used computer adaptive teaching tools.

- Instructional Technology
    - Questions included: type and number of courses that included instructional technology (outside of specific instructional technology courses).

- Candidate Teaching Assessment
    - Questions in this block asked: whether candidate teaching was videotaped and reviewed as well as if candidates received a formal teaching assessment or if it was through courses

    - Additionally, questions asked whether pre-and post-test exams of elementary students were used in candidate teaching assessment.

- Mentors in Student Teaching

    - Questions asked: whether candidates were paired with a mentor during their student teaching experience, and if so if the mentor reviewed the candidate's teaching. Mentors were defined as an affiliated faculty or alumni and were differentiated from the university supervisor.

- Noted Program Attributes
    - Questions in this block asked: what candidate, faculty and survey respondents noted as an aspect that the program does especially well and sets the program apart from others.

In addition to program elements, the survey asked whether programmatic changes occurred during the study period (2014-2018). At the end of each section, I asked whether there were changes in the respective category during the study period and if so, details about the change. These questions were asked in time interval periods such as "from the time period of 2014 and onward, have the required number of course credits changed?" to allow easier recall and comparison from the respondent. Each affirmative response to changes in training approaches were then coded into dichotomous variables (yes=,1 no=0) and were then tested within the statistical analysis to see if the absence of the school or the training element changed the effectiveness of other components.

3.3.5 Student and Teacher Data

Longitudinal data on student achievement scores and their teachers for this study were provided by the NCERDC housed at Duke University (for academic years 2014-2018) (NCERDC, 2019). The NCERDC is a data storage center comprised of data from the North Carolina Department of Public Instruction (NCDPI) as well as the National Center for Education Statistics (NCES). As such, the data within this portal provided a robust portrait of students, teachers, schools, and other aspects of elementary education within the state of North Carolina. This study used the following NCERDC datasets:

School Data:

1) School-Level Report Cards

Student Data

2) Student Course Membership
3) Student Masterbuild File
4) Student Tests

Teacher Data

5) Teacher Education History
6) Teacher Licensure
7) Teacher Pay
8) Teacher Demographics

School Datasets:

- *NC School Report Cards*: The data included within these files have one record for each school and are based on all grades within that school. One file per year was used (2014-

2018), and an average of the last 3 years was used in the analysis. Data from this file that were used in this study include information on district and county of the school, school type (charter, traditional, etc.), number of economically disadvantaged students, number of total students and teachers, number of English as a Second Language students. Additionally, school data on student gender and racial demographics per school were provided directly from the NCDPI website.

Student Data:

- *Student Course Membership:* This file provided information on the course enrollment of each public-school student in North Carolina (for each academic year between 2014-2018). Course information included course code, semester, period, school, district, as well as teacher of record for the course. One record per course is included and therefore multiple student records were in the file. Students and teachers are labeled with a unique identifier so multiple files can be merged for the analysis.

- *Student Masterbuild Codebook:* This file provided data on public school student demographics, school and district of student, attendance records, achievement level, accommodations, and proficiency. There is one record per student in the file (and one file per academic year).

- *Student Test Data:* Test data includes information on all public-school students in grades 3-12 (when testing begins until the last year of testing), as well as the test score, the test ID, test date, and exemption code. The file includes one record per test per student in the file (and one file per academic year for years 2014-2018).

Teacher Data:

- *Teacher Education:* Education data includes the degree of the teacher and the university where the teacher received the degree. Multiple records may occur if a teacher has more than one degree (such as a teacher with a BA and MA degree). The university that the teacher received her degree does not necessarily indicate where the teacher became licensed. There is one file per academic year.

- *Teacher Licensure*: The file includes a record of all certified teachers who were teaching in North Carolina public schools in a given year. License data includes type of license, basis for license, and years of experience. There is one record per license per teacher, and therefore if a teacher had more than one license there would be multiple records of that teacher in the file.

- *Teacher Pay:* This file only included data on certified teachers that are paid by the NCDPI. Data includes teacher salary, school and district of the teacher, education, and years of experience per fiscal year. One record per teacher is included.

- *Teacher Demographics*: The file contains demographic data on all teachers in North Carolina Public schools. The file contains one record per teacher.

Other Data

- *North Carolina Unemployment Data:* For the analysis on teacher attrition, I also included the unemployment rate for counties where teachers taught. This is a similar approach that was used in Goldhaber & Cowan's (2014) study on teacher education and attrition, as they controlled for local county unemployment rates to highlight whether there are other job opportunities available in the area. For this study, I used unemployment data as provided by the North Carolina Department of Labor for years 2014-2018 (and an average employment rate for each county during this period was used in the analysis).

3.4 Building the Data File

The first step in building the dataset for the analysis of student achievement[13] and program elements was to link students with test scores in reading or math to their respective reading or math courses[14] between the years of 2014-2018. This step was necessary to identify the teacher of each student. Courses, and subsequently students, with multiple teachers of record were removed from the dataset. I then linked all teacher files together to identify teachers with 5 years of experience

[13] I built a separate file for each subject, however the steps were identical.

[14] Math and reading courses were identified through their titles, such as "5th grade math". In rare cases, courses labeled as STEM (science, technology, engineering, and math) were the only math course the student was enrolled in, and such was counted for the student's math course. Reading courses were identified as English, reading, or language arts. If a student was in both reading and language arts courses, reading was the selected course.

or less and those that were trained at a program in North Carolina. Additionally, I used both the teacher's type of license (specifically licensed because of BA degree) as well as teacher's education level (BA) to link teachers to their training program. Then the student files and teacher files were linked using the student and teacher id. The sample restriction that had the largest impact on student and teacher sample size in the study was linking students to early career elementary teachers. Only 30 percent of teachers in the initial datafile had 5 or fewer years of experience. Furthermore, while most elementary teachers teach both subjects, for this study more than 70 percent of teachers overlapped. This indicates that some schools have a unique math or reading teacher, and thus some elementary teachers in this study do not teach both subjects.

The datafile for teacher attrition was a separate file than that of the student achievement analysis, with one record of each teacher each year taught, as provided the salary file. Teachers were included in the sample if they had a record in the salary dataset between the years 2014-2018. The focus of this study was to see a teacher's career longevity in their first five years of teaching and were therefore removed if they began teaching before 2009. Variables for a teacher's duration as well as whether the teacher 'quit' (no longer in the file) were then built to be able to perform the analysis on teacher attrition. I then merged this file with student course information (using the de-identified teacher code) to identify elementary teachers (those that taught grades 1-5). These files were then linked to a teacher's training program and only those trained at a program in North Carolina were selected (the process to determine a teacher's program was the same as for student achievement). School and district data were also linked to this file to be added as covariates.

Finally, the data from the document reviews of teacher training programs and program surveys were linked to the training programs (using the IHE id of the university where the program was housed) to test the relationship between these elements, student gains in achievement scores in math and reading, and teacher attrition. Responses from surveys were coded into either dichotomous or continuous variables depending on data collected, survey responses, and type of question. Program element variables were coded the same, regardless of analysis (student achievement and attrition).

3.5 Variables

3.5.1 Dependent Variables

The dependent variable used in the student achievement analysis was elementary student gain scores in math and reading for students taught by entry-level teachers trained in North Carolina over the study period from 2014-2018. Gain scores were built from student end of the year scores, which are taken once per year in the spring semester starting in 3rd grade. While students could have 2 gain scores (as the datafile has student scores for 3rd, 4th and 5th grades), in this dataset they typically have 1 gain score as there are fewer 3rd grade students in the data. Gain scores were used to measure a teacher's effectiveness as they are able to demonstrate the change in a student's achievement in one year. Finally, math and reading were analyzed independently in separate datasets.

For the analysis on teacher attrition, the dependent variable used in this study was a teacher's attrition between the years of 2009-2018. The data used was a long-form database where each teacher had a record for each year that he or she taught. I then calculated a variable for teacher's number of years of experience (or records in the datafile). I also calculated the year the teacher 'quit' teaching, based on the last year the teacher had a record in the datafile. Teachers ‘entered’ the study at any point, as entry was not restricted to 2009. Finally, I used the salary file because the teacher would still be in the salary file if a teacher was on leave or had a temporary absence whereas the teacher would not be in files that only identify teachers of record.

3.5.2 Independent Variables

Since the focus of this study is the effects of program elements, these variables serve as the key independent variables of interest for this study. Program element variables were derived programmatic data from surveys of teacher training programs. Most of these variables were dichotomous such as program requires/has X=1 or program does not require/have X =0. However, there are occasions where the number of courses, or the number of hours required could be the distinction between programs, as such these were also modeled as continuous variables. I identified seven categories of program elements:

1) *Faculty Teaching*: Includes data combined from the US Department of Education and survey responses on the percentage of full-time faculty, percentage of adjunct faculty, number of courses taught by adjunct faculty, and faculty teaching incentives. All variables were continuous, except faculty teaching incentives, which was dichotomous. If a program selected 3 or more identified incentives in the survey drop down, this was coded as 1, all others were coded as 0.

2) *Teaching Experiences Prior to Student Teaching*: Data combined from the US Department of Education, reviews of program curricula and survey responses include: number of courses with active teaching, early-student teaching

opportunity, classroom experience in intro course, computer adaptive student teaching, and number of hours of student teaching. Additionally, dichotomous variables of high number of courses with active teaching (15 or more) and low number of courses (5 or fewer) were added based on high and low quartiles of the continuous variable.

3) *Required Courses:* Courses identified from curriculum reviews were grouped into categories based on their subject matter. The course categories identified were: method courses, subject matter courses, content teaching courses, foundation courses, and diverse learning audience courses. Additionally, data was used from the US Department of Education on the number of hours prior to student teaching. All variables were continuous with the number of courses being the tested value.

4) *Instructional Technology:* Programmatic curricular review and survey responses provided information on the number of courses exclusively on instructional technology, and the number of courses that incorporated instructional technology in the course work. These two variables were continuous with the number of courses being the tested value. Dichotomous variables of high number of courses with instructional technology (10 or more) and low number of courses (3 or fewer) were added based on high and low quartiles of the continuous variable.

5) *Assessment of Candidate Teaching:* Responses from surveys were added to this section to include: the number of courses that required videotaped assignments (of

the candidate's practice), process by which candidate teacher practice (formal process or through courses) as well as whether elementary student pre-post-test scores were included in the assessments. The number of courses that recorded a candidate's teaching was a continuous variable, whereas all others were dichotomous. Additionally, dichotomous variables of high number of courses with active teaching (6 or more) and low number of courses (2 or fewer) were added based on high and low quartiles of the continuous variable.

6) *Mentoring During Student Teaching:* Responses from surveys were added to this section to include whether the program paired candidates with mentors and if so, did mentors formally review the candidate's teaching. Mentors were defined as an affiliated faculty or alumni and were differentiated from the university supervisor. Both variables were dichotomous with yes=1 and no=0.

7) *Notable Program Attributes*: Survey responses to open-ended questions provided four responses to this category: diverse candidates, faculty teaching of candidates, student teaching experience, and candidate preparation for state exams. All were coded as dichotomous variables.

3.5.3 Covariates

Student, teacher/classroom, and school covariates were also included in this study to be able to control for bias that may occur from classroom and school placement as well individual student and teacher differences. These covariates are:

- Student*:*
    - *Race*: dichotomous variable, minority=1 and non-minority (white)=0
    - *Gender*: dichotomous variable, male=1 and female=0
    - Economically Disadvantaged Status: dichotomous variable, yes=1 and no=0
    - *English proficiency*: dichotomous variable, yes=1 and no=0
    - *Year:* Continuous variable of year of score (2015, 2016, 2017, or 2018)

- Classroom
    - *Race*: Continuous variable of percentage of minority students in classroom
    - *Economically Disadvantaged Status*: Continuous variable of percentage of economically disadvantaged students in classroom
    - *English proficiency*: Continuous variable of percentage of English as a Second Language students in classroom
- Teacher
    - *Race:* dichotomous variable, minority=1 and non-minority (white)=0
    - *Gender:* dichotomous variable, male=1 and female=0
    - *MA Degree*: dichotomous variable, has master's degree=1 and does not have master's degree=0
    - *University:* dichotomous variables for each university in the dataset, 1=yes teacher was trained at university 0=no teacher was not trained there
    - *Years of teaching experience*: 1-5 years with year 5 as a reference

- School:
    - *High Percentage of Minority Students:* dichotomous variable for high percentage, yes=1 and no=0. High for this variable is 30 percent or more minority students in the school (this was based off the top quartile distribution of the data).
    - *High Percentage of Economically Disadvantaged Students:* dichotomous variable for high percentage, yes=1 and no=0. High for this variable is 64 percent or more economically disadvantaged students in the school (this was based off the top quartile distribution of the data).
    - *High Percentage of English as a Second Language Students:* dichotomous variable for high percentage, yes=1 and no=0. High for this variable is 57 percent or more English as a Second Language students in the school (this was based off the top quartile distribution of the data)

- County
    - *High Percentage of Unemployment*: For teacher attrition analysis only, dichotomous variable of high unemployment in the county where the teacher's school is located. High for this variable is a county unemployment rate of 4.7 percent or higher (this was based off of the top quartile distribution of the data).

3.6 Descriptive Statistics and Sample

3.6.1 Elementary Students

Table 3.1 below provides demographic data for students in the reading and math samples. After all sample criteria were applied, there were 11,328 students in the reading sample, and 12,605 students in the math sample. Overall, white students were the largest racial category with approximately 40 percent of students, black students were the second largest racial category with almost 30 percent of students in the sample, and Hispanic students were the third largest category, with approximately 20 percent of students. Additionally, more than half of sampled students (59-61 percent) were economically disadvantaged. The means of student gain scores were noticeably different by subject. For instance, the mean of all math gain scores was 0.15 with a standard deviation of 5.75, compared to reading, which was 6.05, and the standard deviation was 6.32. Additionally, the skewness was close to 0 and kurtosis was only slightly above 3. This suggests achievement outcomes for both subjects was normally distributed.

Table 3.1 Descriptive Statistics on Student Sample (by Classroom)

	Reading Mean (SD)	Math Mean (SD)
Male	0.50 (0.31)	0.51 (0.29)
White	0.44 (0.38)	0.40 (0.37)
Black	0.27 (0.32)	0.29 (0.32)
Asian	0.03 (0.10)	0.03 (0.11)
Hispanic	0.21 (0.27)	0.23 (0.28)
American Indian	0.01 (0.07)	0.01 (0.07)
Mixed Race	0.04 (0.12)	0.04 (0.13)
EDS	0.59 (0.34)	0.61 (0.33)
LEP	0.14 (0.23)	0.16 (0.24)
Gifted	0.10 (0.20)	0.09 (0.19)
Disability	0.09 (0.20)	0.10 (0.20)
Gain Score	6.05 (6.32)	0.15 (5.75)
Skewness	0.16	0.04
Kurtosis	3.72	3.40
N	11,328	12,605

EDS= Economically Disadvantaged Student
LEP= Limited English Proficiency

3.6.2 Teachers

Table 3.2 below provides descriptive data on teachers included in the two analyses of this study. Approximately less than 10 percent of sampled teachers had an MA degree in the student achievement analyses, whereas 20 percent had an MA in the attrition analysis. This is to be expected as master’s degrees are fewer among entry level teachers and the attrition analysis included teachers with over 5 year of experience. Furthermore, teachers in the gain score analysis have a mean of 3 years of experience whereas those in the attrition analysis have 5 or more years of experience. Other findings on teachers within the sample show that the teaching population was predominantly white (between 77-84 percent of reading and math teachers), while 12-16 percent of teachers were African American, and approximately 1-2 percent were Hispanic.

Table 3.2. Descriptive Statistics on Teacher Sample (by Classroom)

	Reading Mean (SD)	Math Mean (SD)	Attrition[15] Mean (SD)
Male	0.08 (0.18)	0.12 (0.22)	0.10 (0.20)
White	0.77 (0.33)	0.80 (0.29)	0.84 (0.26)
Black	0.16 (0.28)	0.15 (0.26)	0.12 (0.23)
Asian	0.01 (0.05)	0.01 (0.04)	0.01 (0.05)
Hispanic	0.02 (0.10)	0.01 (0.06)	0.02 (0.09)
American Indian	0.02 (0.10)	0.01 (0.08)	0.01 (0.08)
Other Race	0.03 (0.13)	0.02 (0.12)	0.01 (0.05)
Master's Degree	0.08 (0.20)	0.09 (0.21)	0.22 (0.26)
Traditional License	0.96 (0.14)	0.96 (0.13)	0.93 (0.18)
1 Year of Experience	0.20 (0.28)	0.19 (0.27)	0.03 (0.13)
2 Years of Experience	0.18 (0.26)	0.16 (0.24)	0.04 (0.12)
3 Years of Experience	0.23 (0.29)	0.24 (0.30)	0.12 (0.22)
4 Years of Experience	0.18 (0.26)	0.22 (0.29)	0.22 (0.25)
5 Years of Experience	0.17 (0.27)	0.15 (0.24)	0.16 (0.22)
6 or More Years			0.41 (0.31)
2015	359	401	897
2016	524	580	1,498
2017	661	671	2,138
2018	381	396	8,105

3.6.3 Teacher Training Programs

Table 3.3 below provides number of teachers, per training program that were included in each analysis (after all filters were applied). This table highlights the vast size differences of some of these programs as some programs have less than 10 teachers in the final analytic sample, whereas others have more than 100.

[15] Further information on teachers in the attrition sample such as years of experience are in chapter 5.

Table 3.3. Number of Teachers by University Training Program

University	Math Analysis N	Reading Analysis N	Attrition Analysis N
1	196	182	1,630
2	196	178	1,338
3	128	139	1,033
4	114	104	825
5	88	88	827
6	47	44	357
7	27	31	123
8	70	66	490
9	53	48	366
10	25	26	370
11	17	12	182
12	35	26	211
13	25	30	184
14	24	16	195
15	18	17	164
16	19	16	76
17	17	16	128
18	27	27	187
19	20	17	132
20	18	18	136
21	8	7	76
22	12	13	94
23	7	19	132
24	17	17	94
25	10	10	82
26	15	14	86
27	11	10	119
28	9	10	97
29	10	10	99
30	11	10	87
31	-	-	53
32	-	-	68
33	-	-	63
34	-	-	60
Total	1,282	1,229	10,243

3.6.4 Survey Sample

Programs were selected to receive the survey if they demonstrated a relationship to student achievement gains in my initial analysis (details of this analysis are provided in Chapter 4). I identified 24 programs to receive the survey[16]. At the close of the survey period, 16 out of 24 program coordinators responded, with a response rate of 67 percent. The 24 programs that were contacted to complete the survey trained 83 percent (1,160) of elementary teaching graduates in the academic year of 2016-2017 (US Department of Education, 2016c). Of the five programs that train the most teachers in the state, four responded (57 percent of all elementary education graduates in 2016-17 in North Carolina). Of the 16 schools that responded, 6 were private institutions, while the other 10 were public. Table 3.4 below shows there were 748 elementary math teachers trained by responding programs (and taught 7,724 students), and 741 elementary reading teachers trained by the responding programs (who taught 7,191 students).

Additional demographic data from universities in the survey sample highlight the diverse populations within training programs. Four of the 16 responding institutions trained teacher populations where more than half were minority teachers, and those teachers subsequently taught students who were predominantly minority students. While this is more than in the overall sample, only 20 percent of all teachers in the survey sample were minority teachers, similar to the initial sample of teachers. Also, classrooms had high percentages of economically disadvantaged students and racial minority students, though students with limited English proficiency were less prominent.

---

[16] During the preliminary analysis, 2 institutions showed statistical significance were surveyed and responded to the survey. However, in the final analysis these institutions were no longer significant, but they responded to the survey their responses are included in the analysis, for a total of 24 institutions.

Table 3.4. Descriptive Data of Teachers in Survey Sample by Training Program

	Math						Reading					
	No. TCHR	Non-White Teacher (%)	No. Students	ESL (%)	EDS (%)	Non-White Student (%)	No. TCHR	Non-White Teacher (%)	No. Students	ESL (%)	EDS (%)	Non-White Student (%)
Univ. 1	9	33	65	23	72	65	10	40	73	7	67	64
Univ. 2	25	8	245	19	65	55	30	14	234	11	64	53
Univ. 3	196	11	2,262	15	63	65	182	12	1,842	14	61	65
Univ. 4	24	67	238	14	61	72	16	69	139	25	61	82
Univ. 5	17	6	111	10	55	47	17	6	160	3	56	35
Univ. 6	10	0	53	2	43	19	10	0	88	10	35	31
Univ. 7	27	81	288	17	57	81	27	85	256	15	55	84
Univ.8	47	13	574	16	49	66	44	9	454	15	49	57
Univ. 9	8	13	155	18	70	65	7	14	148	13	64	68
Univ.10	25	56	333	16	62	74	26	38	276	13	63	62
Univ.11	114	21	1,196	18	61	58	104	23	1,180	12	56	55
Univ.12	88	8	850	14	65	59	88	6	830	15	61	54
Univ.13	128	17	1,081	17	52	58	139	21	1,251	15	51	56
Univ.14	18	94	143	23	85	93	18	94	101	19	82	92
Univ.15	-	-	-	-	-	-	10	20	69	9	58	62
Univ.16	12	8	130	20	50	53	13	8	90	18	50	58
Total	748	21%	7,724	16%	60%	63%	741	21%	7,191	14%	57%	60%

EDS= Economically Disadvantaged Student
ESL= English as a Second Language

Of the 8 training programs that did not respond, 5 were public institutions and 3 were small private liberal arts universities with 380 math teachers (teaching 3,786 students), and 362 reading teachers (with 3,266 students). Teachers in both math and reading from non-responding programs were overwhelmingly white. For instance, while 86 percent of teachers were racial minorities in one program in math, for the 7 other non-responding programs, 80 percent of teachers were white teachers (in both subjects). Overall, 9 percent of the reading sample and 8 percent of math teachers of non-responding programs were racial minorities. A similar trend was present with the percentages of minority student taught by teachers from these programs. While 30-50 percent of students taught by graduates of 7 programs were racial minorities, there was1 program (in math) where approximately 90 percent of students taught by their graduates were racial minorities. Teachers from non-responding programs taught similar percentages of economically disadvantaged students as respondents as only 1 program (in both subjects) taught in classroom where less than 40 percent of students identified as low-income. Finally, teachers from non-responding programs taught in classrooms where less than 30 percent of students were English as a Second Language learners.

3.7 Analytical Approach

3.7.1 Survey Design

This study initially drew from qualitative methods using interviews, program document reviews, and surveys to identify program elements of quality among teacher training programs in North Carolina. While reviews of program documents and program curriculums were publicly available, the training methods used within programs were not publicly available. As such, I surveyed programs on training techniques used within their courses as well as assignments, to better understand their training processes and the elements that might set them apart from one another. To design the survey, I reviewed prior literature on teacher training techniques and interviewed program administrators at the three largest elementary education training programs in North Carolina. Program size was based on the number of graduates from programs in elementary education, elementary math, and elementary reading as reported by the NCDPI in 2017. Combined, 30 percent of teachers trained in North Carolina in 2017 graduated from these programs (US Department of Education, 2016c). In total, I interviewed four program directors as one of the three programs had co-directors.

Interview questions were designed to understand program culture, reform efforts of the programs, and training techniques. Interviews ended with open-ended questions to provide room for interviewees to share additional information about programs (see Appendix B for interview protocol). The protocol and questions were designed based on the research questions guiding the study, as well as insight from previous literature on training components as noted in Henry, Campbell et al., (2013) and Preston (2017). Interviews were recorded with the Tape A Call App and transcription services were provided by Google YouTube transcription services. I first reviewed the interview transcript along with my own notes and added additional notes where necessary. I then sent the transcript (with notes) to interview participants for their feedback. Once feedback and refinements were received, I made a list of each question asked and respondent answers. From there, I highlighted commonalities and differences for each question.

The five themes I structured the interviews around were: faculty qualifications and attrition, teacher candidate assessment and feedback, presence of mentors during the candidate's student teaching experience (in addition to the cooperating/supervising teacher), data used to inform teacher training (such as edTPA or evidence of student learning), and the use and training of instructional technology. Responses from these interviews provided a foundational understanding of components of teacher training in North Carolina. Furthermore, they also reiterated broader themes within the state such as the increased use of edTPA as a training and assessment tool as well as an overall emphasis on program standardization within the state.

3.7.2 Survey of Training Programs

The interviews highlighted several areas previously identified that were relevant within teacher training. From those insights and the vast literature on teacher training, I drafted the survey with the following sections: faculty teaching, teaching experiences prior to student teaching, instructional technology, assessment of candidate teaching, mentoring during student teaching, notable program attributes, and a timeline for changes in program offerings (for those that occurred between 2014-2018). The survey was designed using Qualtrics software and was emailed to program coordinators using a personalized link to allow for only one response per program. The survey was in the field from December 13-February 10th. The survey was open for a longer period due to the academic calendar of most universities in this sample, as most campuses were closed or faculty were on vacation. For this reason, follow-ups and survey reminders were not sent between December 20-January 5th. While the survey was still open during that period, I assumed most program representatives would not be available to complete the survey given university closures and did not want to unnecessarily contact programs.

To increase the response rate, four different techniques were applied. The first was a financial incentive of $25 for every respondent. Additionally, the survey length was designed to assist with response rate. Although there were 58 questions, I reduced the number of open-ended responses, and the estimated completion time was between 10-15 minutes. I also added skip patterns to assist with the speed of answering questions if respondents did not know the answer to a question. Finally, as this was a survey sent through email, I sent email reminders to each program 5 times and then I personally called those that had not responded before the survey closed (Mertens, 2005). For this study, calling survey respondents had the largest impact on the response rate. Before calling each representative, only 9 coordinators had responded. However, after calling the remaining 15 representatives, 7 coordinators responded within 24 hours.

Although it is difficult to discern why program coordinators would not respond to the survey, insights from the preliminary interviews and responses to open-ended questions in the survey shed light on possible reasons for a non-response from programs. The first possible reason is that programs have experienced increased pressure from the state of North Carolina to revise training program requirements. As such, programs may be sensitive to answering questions about their programs. Another possible reason for non-response could be tied to motivation. While there was a $25 incentive for those that completed the survey, three respondents chose to not receive the incentive, and 6 programs noted in email communication or in open-ended questions that they only responded because they wanted to assist with the research process of a graduate student. Additional possibilities for non-response could be due to components of the survey itself. For instance, the survey was initially administered during a period when institutions of higher education were known to be closed or when staff were on vacation, as such representatives may not have been available to complete a survey. Another possible rationale is that the survey itself has 58 questions with several open-ended questions. Although respondents did not take long to complete the survey (and open-ended questions were requests for short answers such as course titles), when initially reviewing the survey, respondents may have assumed the survey would take too long to complete.

3.7.3 Coding of Survey and Program Data

After the survey closed, I analyzed which questions elicited responses that could be used in the quantitative analysis. First, I examined responses of each question to determine if responses matched intended questions. For instance, when I asked if faculty teaching quality was included in faculty reviews, every respondent stated, yes faculty teaching was reviewed by students in end of the semester surveys. However, I had sought to understand if faculty teaching was reviewed using other means (given the literature is mixed on the effectiveness of student reviews of faculty teaching) (Braga, Paccagnella, & Pellizzari, 2014). As such, I removed this question from the analysis. I also examined survey responses to see if there was sufficient variation in responses (Nardi, 2018). For some questions such as what type of training course record candidate training, every program noted candidates were videotaped in method courses. Given there was no variation, I removed this question in the latter analysis.

After each categorical question was analyzed, I examined open-ended questions. Open-ended questions in this survey were used for 2 purposes: to cross check answers with specific responses (such as course number and title) and to provide context (Engle, Jann, Lynn, Scherpenzeel, & Sturgis, 2014). For instance, some programs said students had their first teaching experience in their first training course. By using the course number provided, then verified this information using course catalog data. Furthermore, I asked when the program most recently modified a specific training element. Open-ended questions highlighted that program changes (such as candidate assessment and course sequencing) was primarily due to the implementation of the state required assessment tool edTPA. As these changes were representative of state requirements and not programmatic changes, I did not include them in my analysis. Finally, I coded each useable question into variables on elements within each surveyed program. Coding of responses was dependent on the question type. For instance, yes/no questions were coded into dichotomous variables, whereas number of courses with specific elements (such as those that record candidate teaching) were coded into continuous variables. These variables were then used to answer research questions on relationships between training elements, student achievement, and teacher attrition as described below.

3.7.4 Multilevel and School Fixed Effects Models

The first equation used to examine the effects of program elements on student achievement was a covariate adjusted multilevel analysis (Bryk & Raudenbush, 1992; Luke, 2004). There are two-levels in the model used for this analysis. The first level is individual students nested within teacher classrooms, and the second level is the teacher-classroom level. This model also clustered standard errors at the classroom level (Cameron & Miller, 2015; Goldhaber et al, 2013; Koedel, et al, 2012). This equation is written as:

$$\boldsymbol{Level\ 1}: \boldsymbol{Y_{icjst}} = \boldsymbol{\beta_{0c}} + \boldsymbol{\beta_1 Y_{icjs(t-1)}} + \boldsymbol{\beta_2 X_{ijst}} + \boldsymbol{r_{icjst}}$$

$$\boldsymbol{Level\ 2}: \boldsymbol{\beta_{0c}} = \boldsymbol{\gamma_{00}} + \boldsymbol{\gamma_{01} P_{pjcst}} + \boldsymbol{\gamma_{02} C_{cst}} + \boldsymbol{\gamma_{03} J_{jsct}} + \boldsymbol{\mu_{jpcst}}$$

Equation 3

In this model, Y_{icjst} is the student achievement score for student *i* in classroom *c*, for teacher *j*, in school *s*, at time *t*.

β_{0c} represents the model intercept value the classroom's *j*'s average score at time *t-1*

γ_{00} represents the mean value of the level 1 dependent variable (the student) controlling for the level 2 predictor (teacher classroom)

$Y_{icjs(t-1)}$ is student *i*'s lagged test score from the previous year,

X_{ijst} a vector of characteristics of student *I*, for teacher *j*, in school *s*, at time *t*,

P_{pjcst} (the key variables of interest) is a vector of elements within teacher preparation programs for program *p* for teacher *j* in classroom *c* at school *s* at time *t*,

C_{cst} is a vector of classroom characteristics for classroom *c* in school *s* at time *t*,

J_{jst} is a vector of teacher characteristics for teacher *j* in school *s* at time *t*,

r_{icjst} represents a student-specific random error term,

μ_{jpcst} represents a classroom-specific random error term.

This multilevel model served as an exploratory model to consider the relationships of student achievement and teacher training within classrooms as well as between classrooms. In addition to the multilevel model noted above, I used a school fixed effects model that examined variation in teacher training elements within schools. The following linear models with the school fixed effect applied is written as:

$$\boldsymbol{Y_{ijcst} = \beta_0 + \beta_1 Y_{ijcs(t-1)} + \gamma_1 P_{pjcst} + \gamma_2 X_{ijst} + \gamma_3 C_{cst} + \gamma_4 S_{st} + \gamma_5 J_{jt} + \phi_s + \varepsilon_{ijpcst}}$$

Equation 4

In this equation, Y_{ijcst} represents student *i*'s test score taught by teacher *j* in classroom *c* at school *s* at time *t*,

$Y_{ijcst(t-1)}$ represents student i's test score in the previous year,

P_{pjcst} (the key variables of interest) represents a vector of elements within teacher preparation programs for program *p* for teacher *j* in classroom *c* at school *s* at time *t*,

X_{ijst} is a vector of characteristics for student *i* taught by teacher *j* at school *s*, at time *t*,

C_{cst} is a vector of classroom characteristic for classroom *c* in school *s* at time *t*,

S_{st} is a vector of school characteristics for school *s* at time *t*,

J_{jst} is a vector of teacher characteristics for teacher *j* in school *s* at time *t*,

Φ_s represents a school fixed effect,

Φ_i represents a student fixed effect,

ε_{ijpcst} is a vector of a random error term.

This school effects model was used to address potential bias from student sorting into schools by estimating the contribution of training programs by comparing the differences of student achievement within the same school across years. While additional biases still exist in the model from teacher placement and university teacher training programs, sufficient variation was not present in the data to explore other fixed models (such as student, teacher, or university). However, to account for this, I added teacher and student characteristics as control variables as these may influence which programs teachers attended as well as the student achievement scores themselves. Additionally, while data limitations precluded me from exploring a university fixed effects model, I added university covariates to examine relationships between the programs themselves, student achievement, and teacher attrition. Finally, the year of the student's score was added as a control to adjust for year-to-year differences due to esoteric changes in test items (Bastian & Henry, 2015; Goldhaber, Gross, & Player, 2011; Henry, et al., 2011; Henry, Purtell, et al., 2014; Koedel, et al, 2012).

3.7.5 Survival Analysis

The third question of this study, effects of training characteristics on teacher attrition, was answered using a discrete time hazard model of attrition as outlined in Kelly & Northrop's 2015 study on college selectivity, teacher satisfaction, and teacher attrition. Using this model, the authors measured the probability of a teacher's attrition given the selectivity of the university the teacher attended, the teacher's preparedness, and his/her satisfaction with teaching overall. This model built upon the Kelly & Northrop (2015) model by adding a control for program characteristics as well as the program where the teacher was trained. The data available on teacher attrition in this study only notes if the teacher is no longer in the data file and thus if the teacher still teaches in a public school within the state of North Carolina. As such this question defined attrition through dichotomous variables where 1= attrition in a public elementary school in the state to the next year or 0=did not continue to the next year. After creating a long form data base to examine the data, a discrete time hazard regression was used to measure the probability that a teacher left the public-school system. The baseline equation used for this analysis was:

$$\mathbf{h_j=Pr[T=j \mid T>=j]}$$

Equation 5

In this model, h_j is the probability that teacher j will leave the profession at time T (Singer & Willet,1993). However, this model only predicts the time at which a teacher will (or will not leave) the public system, without further information about the teacher. As such, the model used applied predictors to refine the estimated probability that a teacher would not leave if she were trained with specific elements. This model then became:

$$\mathbf{h_{ij}=Pr\{T_i=j \mid T_i >=j, Z_{1ij} = z_{1ij}, Z_{2ij} = z_{2ij}, Z_{3ij} = z_{3ij}, Z_{4ij} = z_{4ij}\}}$$

Equation 6

Similar to Singer & Willett (1993) the subscript of is *i* is added to the original predictor definition to separate the teacher and his/her predictor values in a given time period T (assuming he/she continued in the system for the periods prior). Additionally, the following vectors of predictors were added to model: teacher demographic characteristics (Z_{1ij}), demographics of the school where the teacher taught (Z_{2ij}), and program training elements (Z_{3ij}) on teacher (*i*). Similar to Goldhaber & Cowan's (2014) model on teacher training and teacher attrition, this model also included a vector of local employment rates (Z_{4ij}) as employment opportunities (or lack thereof) may also influence whether a teacher leaves the public school system (Adnot, Dee, Katz, & Wyckoff, 2017; Goldhaber, Krieg, & Theobald, 2016; Kelly & Northrop, 2015; Ronfeldt, Lankford, Loeb, & Wyckoff, 2013; Singer & Willett, 1993).

4.0 Program Elements and Student Achievement

This chapter provides results for the first two research questions of this study which examined program elements present in teacher training programs in the state of North Carolina as well as the statistical relationships between those elements and student achievement gains in reading and math. This chapter is organized by research question, as section 1 provides the research findings to question 1 and likewise, section 2 provides results for question two. The research questions answered in this chapter are the following:

1) What program elements, beyond those available in the public domain, are present in teacher preparation programs and are indicators of program quality for elementary teaching programs in North Carolina?
2) What is the relationship between elements of an elementary teacher preparation program and elementary student achievement in North Carolina?

The beginning of each section in this chapter outlines the basic methodology employed to answer each question and is followed by a detailed description of the findings. The first section of this chapter provides results found from interviews, document reviews, and survey responses that were used to understand what program elements are present in training programs in North Carolina. The second section of this chapter provides results from statistical analyses used to understand the relationships between program elements and student achievement are provided detail in the second section of this chapter. This study used a two-level multilevel statistical model as well as school fixed effects model to test these relationships.

4.1 Program Elements

4.1.1 Findings from Interviews

In late 2018, I conducted interviews with program coordinators from the 3 largest teacher training programs in North Carolina. These 3 programs graduated 543 elementary teachers in 2016-2017 (approximately 34 percent of elementary teachers prepared by institutions of higher education in North Carolina that year). These interviews complimented the prominent literature on teacher training and provided a deeper understanding of training approaches specific to elementary teaching candidates in North Carolina. Through this process, I identified 5 overarching components of teacher training programs that could be collected either through program document reviews or surveys of individual programs. These components were: faculty type and incentives for faculty teaching, candidate teaching experiences, use of instructional technology, assessment of candidate teaching, and the presence of mentors during the candidate's student teaching experience.

In addition to the five areas noted above, two broad themes emerged: changes in teacher training due to state regulations and the difficulty of defining differences between programs. This study occurred at the same time that greater standardization in teacher training was occurring both nationally and especially in North Carolina. Interviewees often mentioned state reforms to teacher training in their responses and all discussed the impacts/changes resulting from implementing edTPA[17]. Because of these state-wide requirements and reforms, I included a question about

[17] The state has required all teacher training programs to implement and use edTPA as an assessment of candidates by the academic year 2019-2020.

timing of program-specific initiatives to address the time lag between the data used for this study and changes within programs.

Interviews also highlighted the difficulties in defining instructional approaches used within their own courses and the programs themselves. For instance, only one interviewee was able to answer what teaching techniques faculty members employ (such as lecture, read and respond, microteaching, fishbowl, or other), noting methods depended on faculty member preferences. Furthermore, when programs were asked about course assignments or other training techniques, answers were often undefined and not detailed. However, when questions were specific and asked about faculty classification for those that teach elementary education candidates (full-time faculty or adjuncts) or number of courses that record a candidate's teaching practices, interview responses were more specific and provided detailed insight into elements of training programs in North Carolina. Additional insights from interviewees that were used to guide the survey design and my overall understanding of training in North Carolina are outlined in methods section of this dissertation.

4.1.2 Findings from Program Document Reviews

Responses from the above interviews confirmed much of the literature pertaining to training elements of programs and provided contextual information. However, they also highlighted the difficulty in distinguishing elements between programs. To address this and to further identify elements of programs that may have a relationship with teacher effectiveness and attrition, I collected multiple forms of data on elements within teacher training programs in North Carolina. The first step was to gather publicly available data specific to each program. Through a document review of program course catalogs and Title II reports to the US Department of

Education (2016c), I identified 16 program elements that could influence a candidate's teaching and showed enough differentiation to be tested in a statistical model. The program elements I identified and collected are listed below in Table 4.1, separated by subject (math and reading) and analysis (student achievement or attrition).

Findings from this review of program documents showed that differentiation was greatest among number of hours before student teaching and faculty classifications for those that supervise the student teaching experience. For instance, while the mean number of hours before the official student teaching experience was approximately 146 hours, the standard deviation of 72-74 shows the number of hours were wide spread. This is even more pronounced with the ratio of students to adjunct faculty as the standard deviation is between 25-26, compared to the mean of 15. Although less pronounced than student teaching requirements, there are distinguishable differences in some of the required course categories. For instance, as shown in Table 4.1 courses focused on teaching diverse audiences saw the largest differences among required courses, with a standard deviation of approximately 1.00 and means of 2.33-2.35. One final observation is that although there was limited differentiation among programs and the number of required courses of instructional technology (with a mean of approximately 0.83 courses and a standard deviation of 0.65-0.66), other researchers have found a significant relationship to a teacher's effectiveness, and it is therefore included (Henry, Campbell et al., 2013).

Although there were notable differences amongst programs, there were areas where programs had almost no distinguishable differences. For instance, when examining admission criteria for programs, as well as total number of hours required for the training programs overall, most programs required a 2.7 GPA to apply to the teacher education program and required approximately 120 credit hours to graduate. Another area that showed limited variability is that all programs had a student teaching seminar course at the end of the training program. Although the structure varied somewhat (with some programs having seminar and a teaching methods component at the same time), many courses referred to student teacher experience as seminar, whereas others referred to seminar as a course to guide the student teaching experience. Course descriptions as well as other program sources lacked descriptions to accurately code these requirements, and I therefore excluded these from the analysis.

Findings from this content analysis of public documents provided useful information on the elements that could be examined in the subsequent analysis on program elements. Indeed, this data collection provided information on training elements which had not been previously collected on elementary training programs. However, experts have argued that much of what may distinguish programs is not the varying course offerings, but more the unique programmatic structures, assignments, and techniques within individual programs that are not available through public sources (Preston, 2017). To account for this, I surveyed programs to further gather information on program elements that are not otherwise available. As such, elements of program quality were derived from both this review of program documents and surveys of programs.

Table 4.1. Program Elements from Program Document Reviews

Training Element	Math Mean (SD)	Reading and Attrition Mean (SD)
Institution Selectivity		
Institutional SAT	1029.48 (125.26)	1024.29 (125.11)
Number of Required Courses		
Teaching Methods	1.43 (0.73)	1.42 (0.72)
Math Content	1.87 (0.76)	1.88 (0.74)
Math Methods	1.74 (0.81)	1.75 (0.79)
Reading Content	2.09 (0.90)	2.08 (0.88)
Reading Methods	2.30 (0.97)	2.38 (1.01)
Teaching Diverse Audiences	2.35 (1.03)	2.33 (1.01)
Foundations of Education	2.13 (0.81)	2.13 (0.80)
Assessment	1.00 (0.80)	1.00 (0.78)
Instruction Technology	0.83 (0.65)	0.83 (0.64)
Student Teaching Requirements		
Early In-service	0.48 (0.51)	0.50 (0.51)
Hours Before Student Teaching	146.22 (74.01)	146.38 (72.39)
Hours of Student Teaching	558.52 (89.27)	561.92 (88.88)
University Supervisors by Type		
Ratio of Students to Full-Time Faculty	14.78 (26.46)	14.83 (25.88)
Ratio of Students to Adjunct Faculty	8.78 (11.71)	8.46 (11.56)
Ratio of Full-time Faculty to Adjunct Faculty	4.59 (5.04)	5.30 (6.02)

4.1.3 Findings from Survey on Teacher Training Practices

The survey on program elements was administered to 24 program coordinators at elementary education programs in North Carolina. The programs that were selected to receive the survey were based on an initial statistical analysis of programs (which is outlined in section 2 of this chapter). Once I identified programs to be surveyed, I emailed the survey to the program coordinators as listed on the websites of each program. I received 16 responses out of 24, for response rate of 67 percent. The survey asked 58 questions (including background and information to receive the incentive) and was divided into the following sections: Faculty Teaching, Teaching Experiences Prior to Official Student Teaching, Instructional Technology, Candidate Assessment, and Mentoring of Student Teachers. All responses used in the statistical analysis of student achievement and teacher attrition are provided in Table 4.2 below.

The initial section of the survey, *Faculty Teaching*, asked respondents about the percent of adjunct faculty, number of courses adjunct faculty taught, incentives for improving faculty teaching, and reviews of faculty teaching. The first question asked what percentage of faculty in their department were classified as adjunct, and a follow-up question asked which courses adjunct faculty taught (by giving options such as method only, content only, etc.). Responses indicated that the mean percent of adjuncts that taught candidates was 0.20 (0.12). As for types of courses taught by adjuncts, most respondents selected multiple categories ranging from foundations in education, content only courses, teaching content, teaching methods, and seminar/practicum. Responses were thus made into a numerical variable, as I was able to identify the number of courses that adjunct faculty taught based on data from the aforementioned document reviews. For instance, if a respondent noted adjunct faculty only taught math content courses, I could identify from the review of program requirements that the program required 2 math content courses and thus adjunct faculty taught 2 courses. Findings showed that approximately 8.60 (SD=5.47) courses were taught by adjunct faculty in the math sample (and similar findings in the reading/attrition sample).

Respondents were also asked about incentives for innovative teaching of faculty as well as whether faculty teaching quality was a component of faculty reviews. Questions regarding incentives for teaching had both an open response category and a selection of types of incentives such as: financial awards for faculty teaching, department workshops on teaching techniques, faculty meetings to discuss teaching strategies, support to attend professional associations specifically on faculty teaching. Most respondents selected support for professional meetings on faculty teaching. However, this signaled a potential weakness in the question as I did not ask if faculty actually attended those meetings and the respondent may also have read this question as support to attend conferences about teaching in general instead of conferences focused on a faculty member's own teaching practice. To account for this, answers were made into dichotomous variable where 1 indicated the respondent selected 3 or more categories of support for teaching. The mean was 0.47 (0.52) for respondents in the math sample and 0.44 (0.51) in reading and attrition samples for those that selected 3 or more categories.

The last question in this section did not garner a response that could be used in the survey. This question asked if the faculty teaching quality was used in faculty reviews. Overwhelmingly respondents said yes, faculty are reviewed every semester by their students in faculty reviews. This question sought to understand if faculty teaching was measured in ways other than student surveys, as student surveys have shown mixed results in their efficacy to assess teaching quality (Braga, Paccagnella, & Pellizzari, 2014). However, this question was not worded in a way that would elicit those responses and therefore responses were omitted.

Questions in the second section of the survey, *Teaching Experiences Prior to Student Teaching*, sought to understand a candidate's first classroom experience as well as opportunities for active teaching throughout the program. The first 2 questions asked when a candidate had his/her first classroom experience (this could be observation or informal teaching), and asked which specific course included the experience. Although all respondents selected 'second year', the course where candidates have this experience saw more variation. Findings showed that between 73-75 percent of sampled programs had an in-classroom experience in their first education course. The third question in this section sought to identify how many courses within the program had an active teaching component for candidates (not including the student teaching experience).

The mean number of courses with active teaching experiences was approximately 10-11 courses in the curriculum with an active teaching component (and for both samples the standard deviation is approximately 4). An additional question in this category pertained to computer adaptive teaching experiences. In preliminary interviews, programs that were in rural areas noted they used computer adaptive teaching models as there were fewer in-classroom experiences available.

Table 4.2 below shows that less than half of programs had a computer adaptive experience available for their candidates (the mean of 0.38 SD=0.50 in reading/attrition samples and 0.40 SD=0.51 in the math sample).

The third survey category was *Instructional Technology*. Initial questions in this section asked how many courses in the curriculum had an instructional technology component as well as which courses included technology training (such as method courses, introductory courses, etc.). Although the content review analysis showed that most programs had a course titled “instructional technology”, course descriptions of other required courses did not indicate if there were instructional technology components to other courses. Respondents noted technology training was prevalent throughout their courses as the mean number of courses with a technology component was approximately 5 courses (with a standard deviation between 2.67-2.76) courses).

The fourth section of the survey asked program coordinators about *Assessment of Candidate Teaching* and how programs provided candidates with feedback on their teaching. The first question in this section asked how candidate teaching is reviewed. In preliminary interviews programs noted candidate teaching was videotaped and then programs reviewed their teaching. To understand the prevalence of this technique, I asked whether the program videotaped candidates, and if so, in which courses is it used. All but one program responded that they require candidates to videotape their teaching. While there was not enough differentiation amongst programs to model this question, responses varied on the number of courses that required videotaping. The mean number of courses that required candidates to videotape their teaching was approximately 3 courses with a standard deviation between 2.56-2.61

To understand how candidate's received feedback on their teaching, I asked whether there was a formal review process or if reviews occurred in courses. Most programs noted that feedback was given through courses and seminar and there was no separate assessment (47-50 percent of programs reviewed candidates in courses). Comparatively, the mean for those that provided a separate review was 0.25 SD=0.45 for reading/attrition samples and 0.27 SD=0.46 for those in the math sample. The last question on candidate assessment was to gauge if programs used evidence of student learning (that is learning of students taught by the candidate) in candidate teaching assessments. Although answers varied as to what type of evidence they used (if it was used), 40 percent of programs in the math sample and 44 percent of those in the reading/attrition sample used student pre-and post-test scores in the candidate's assessment.

Section five of this survey asked whether the program used *Mentors*, separate from the coordinating teacher and university supervisor, in the student teaching experience, and whether mentors assessed a candidate's teaching practice. Mentors were defined as an affiliated faculty or alumni who were assigned to provide support to the candidate during their student teaching. Responses highlighted in Table 4.2 show that approximately 53 percent of programs in the math sample and 56 in the reading/attrition samples pair a mentor with a student teacher and approximately 40 percent in math and 44 in reading/attrition samples use mentors to assess the candidate's teaching.

Three questions that were asked before the closure of the survey were *Noted Program Attributes* which asked what attributes graduates, faculty, and program coordinators cited as an area the program does well. Overwhelmingly, respondents answered that graduates cited their student teaching experience as the program’s best attribute (0.73 in math and 0.69 in reading/attrition), however they noted that relationships with faculty was something candidates also appreciated (0.27 in math and 0.31 in reading/attrition). As for what faculty cited as something the program did well, each program had a unique response that varied from preparing candidates for state assessments and licensure exams, to supporting students throughout the experience, and they also noted candidate field experiences. However, 20 percent of those in the math sample and 25 percent of those in the reading/attrition samples noted they prepared candidates well for state exams. Finally, program coordinators responses varied considerably from uniqueness of the community (and diversity of candidates), to student teaching experiences. However, many noted unique and diverse candidates was one thing that set their program apart from others (40 percent in the math sample and 38 percent in the reading/attrition sample).

Given that programs are dynamic and are evolving due to new research, a changing student population, and state regulations, the last section asked if programs had changed aspects of their program over the survey period (between 2014-2018). If programs indicated there was a change, a follow-up question asked the year the change occurred and the specific change. However, many responses were about the implementation of edTPA. For instance, programs noted they changed candidate assessment to incorporate edTPA, or they had changed course sequencing and added edTPA preparation courses to the curriculum. While these indicators of program changes were intended to be added to the model, it was noted that most changes were due to edTPA and were not indicators of programmatic change. Furthermore, though the increased use of edTPA throughout North Carolina was noted within all survey responses, there has been significant research on the use and validity of edTPA recently. For instance, as Gitomer, Martinez, Battey, &

Hyland (2019) found there have been substantial impacts on programs as well as the candidates themselves during the implementation process. These concerns ranged from additional training for preparation programs as well as the validity of the tool to score a candidate's practice. Given the robust literature on edTPA, and the focus of this study on unique program elements, I did not incorporate changes due to edTPA in the analysis.

Table 4.2. Program Elements from Survey Responses

	Math[18] Sample Mean (SD)	Reading and Attrition Sample Mean (SD)
Faculty Teaching		
Percentage of Adjunct Faculty	0.18 (0.11)	0.20 (0.12)
Adjunct Teach all Course Types	8.60 (5.47)	8.38 (5.37)
Faculty Incentives for Teaching	0.47 (0.52)	0.44 (0.51)
Teaching Experiences Prior to Student Teaching		
Classroom Experience in Intro	0.73 (0.46)	0.75 (0.45)
Active Teaching Courses	10.33 (3.52)	10.00 (3.65)
Computer Adaptive Teaching	0.40 (0.51)	0.38 (0.5)
Instructional Technology		
Technology Use in Courses	5.80 (2.76)	5.75 (2.67)
Candidate Assessment		
Videotape Teaching in Courses	3.33 (2.61)	3.44 (2.56)
Formal Assessment	0.27 (0.46)	0.25 (0.45)
Course Based Assessment	0.47 (0.52)	0.50 (0.52)
Pre-Post Student Tests	0.40 (0.51)	0.44 (0.51)
Mentors		
Mentors during In-Service Training	0.53 (0.52)	0.56 (0.51)
Mentor Review of Teaching Practice	0.40 (0.51)	0.44 (0.51)
Noted Program Attributes		
Student Teaching Experience	0.73 (0.46)	0.69 (0.48)
Faculty	0.27 (0.46)	0.31 (0.48)
Preparedness for State Exams	0.20 (0.41)	0.25 (0.45)
Diversity	0.40 (0.51)	0.38 (0.50)

[18] 15 of the 16 respondent institutions were in the math sample, whereas all 16 respondents were in the reading and attrition samples.

4.2 Program Elements and Student Achievement

4.2.1 Multilevel Models and School Fixed Effects Analysis

After identifying program elements from document reviews as well as from surveys of the programs themselves, I employed multiple analytical processes to examine the relationship between these elements and student achievement gains. The first step was to examine results from a baseline multilevel model and a baseline school fixed effects model to understand the interaction between student and classroom/teacher covariates in both subjects. Then, using a multilevel and school fixed effects model, I examined the relationship between student gain scores and the university where the teacher received his/her training. This analysis also identified which programs to survey. After these preliminary analyses, I examined the relationship between program elements and student gain scores themselves. To fully investigate these relationships, I tested each program element individually as well as in blocks by topic in a multilevel model (with student, classroom and teacher covariates) to see if the element itself had any effect on student gains. This process helped identify variables with a significant relationship to student achievement gains. I then identified variables to enter into a full multilevel and full school fixed effects model to test the extent of relationships with student achievement gains. Each step listed herein, and the subsequent results are described in detail below.

4.2.2 Baseline Model

As discussed in the methods section, due to the nesting structure of students within classrooms and students within schools, fundamental assumptions for an OLS regression were violated and therefore multilevel models as well as fixed effects models were used in this analysis (Garson, 2020). Initial multilevel analyses are provided below by subject (Table 4.3 for models 1-5 of reading, and Table 4.4 for models 1-5 of math). Model 1, the baseline model, tested whether the teacher-classroom (level 2) had a significant effect on the student (level 1), without any covariates (note, in the tables level 2 is labeled simply as the teacher level).

The second model tested the relationship when the following student covariates were added: economically disadvantaged status, racial minority, male students, ESL students, and a variable for the actual year (to control for year to year differences). The third model further examined this relationship when classroom and teacher covariates were added alongside student covariates from model 2. The classroom covariates included were: percentage of economically disadvantaged students, percentage of English as a Second Language students, and percentage of minority students in the classroom. Teacher covariates added were: male teachers, minority teachers, master's degrees, and number of years of teaching experience (1^{st},2^{nd},3^{rd}, or 4^{th}; the 5^{th} year was omitted).

The final models examined in this section were school fixed effects models that examined the variation of teachers within schools in a given year (shown in Table 4.5 for both subjects). Model 4 was a baseline model with only school fixed effects and the student's gain score, whereas model 5 included student and teacher classroom covariates. For both models, school fixed effects were added using the school id where the student attended. A school fixed effect was added to this analysis to further examine within school variation, given that a fixed effect increases the inferential power by reducing the variation within the grouping variables (in this case the school). As each school serves as its own control in these models, bias was reduced in this model due to omitted school-level covariates, and thus provided additional information on relationships between a teacher's training and achievement gains of his/her students that were otherwise not seen in the multilevel models (Rabe-Hesketh & Skrondal, 2012).

Before examining the relationships between student gains and covariates, I examined the attributes of the baseline multilevel model, Model 1. The variance partition (the intraclass correlation or ICC) for the baseline model was 0.09 for reading and 0.18 for math, indicating 9 percent of the variance in reading student gain scores was between students' classrooms (and 18 percent in math gain scores was from the student's classroom). Furthermore, the random effect parameters showed the classroom variance was 3.75 in reading and 5.91 in math. However, the large residual variance in both subjects (36.25 in reading and 27.43 in math) suggests much of the within classroom variance in gain scores was still unexplained even after controlling for the classroom effect (Garson, 2020). A comparison of the variance for each of the initial multilevel models are shown at the bottom of Table 4.3 (for reading) and Table 4.4 (for math).

When student covariates were added to the baseline model, the variance for the classroom as well as the residual variance did not change substantially in reading or math. For instance, when examining the total variance explained by covariates, for model 2 this was 0.00 for math and 0.01 in reading, and in model 3 it was 0.01 for reading and 0.02 in math. This indicates that almost none of the variance was explained by covariates. To further examine the variance between these models, I examined the proportion of student and teacher-classroom variance that was explained by the covariates. Compared to the baseline model, covariates that were added to model 3 explained 5 percent of teacher-classroom variance for reading and 7 percent for math. However, none of the student variance was explained by covariates in model 3 for either subject.

These initial models showed a few significant variables in both subjects. For student level covariates, the first that showed a strong positive relationship to gains in both reading and math, was a student's English proficiency, with a coefficient of 0.81 (SE=0.18) in reading and 0.73 (SE=0.16) in math and was significant at the $p<0.001$ level for both subjects in model 2. Similar findings were also present in model 3, at the same significance level. This suggests that students who are English language learners have almost a full point increase in gains in either subject compared to those who are already proficient in English. Additionally, the year of the gain score (such as 2015) was significant. In model 2, for reading the coefficient was 0.35 (SE=0.07), significant at the $p<0.001$ level. For math, while significant at a similar level, the coefficient in model 2 was -0.28 (SE=0.07). Finally, for reading, there was a negative relationship between gain scores and minority students in model 2 with a coefficient of -0.34 (SE=0.13) at the $p<0.05$ level (though not significant in model 3).

For teacher-classroom and teacher covariates, the first that showed a significant relationship in both reading and math was teacher experience, relative to 5th year teachers. For instance, in reading, teachers with one year of experience had a significant negative relationship to student gain scores with a coefficient of -0.96 (SE=0.28) at the $p<0.001$ level (shown in Table 4.3, model 3). This relationship continuously became less negative (though still significant) for each of the following years, with a coefficient of -0.60 (SE=0.23, $p<0.05$) in the teacher's fourth year. This trend is similar in math as the coefficient for first year math teachers was -1.51 (SE=0.28, at the $p<0.001$ level). For second year math teachers, this slightly declines to -1.03 (SE=0.26 at the $p<0.001$ level) (shown in Table 4.4, model 3). By year three the coefficient is no longer significant, though the relationship was still negative at -0.15 SE=0.25 (and almost positive by year four with a coefficient of -0.08 (SE=0.22), though not significant). These findings suggest two things: students taught by entry level teachers have lower gains in reading and math scores than those taught by teachers with more years of experience[19]. However, it also shows that this negative effect lessens by each year of experience, especially in math. Such results align with the literature on teacher effectiveness as student scores have been shown to be higher for teachers with more experience. Additionally, this literature has been the basis for much of the research in teacher training as researchers investigate training methods that influence a teacher's effectiveness in the initial year as the teacher of record (Rice, 2010).

One additional teacher covariate that had a significant negative relationship to student gains for math teachers, was a master's degree (Table 4.4). This had a significant negative relationship with a coefficient of -0.86 (SE=0.26) at the $p<0.001$ level. This suggests that gains in student math scores for students taught by a teacher with a master's degree were almost one point lower than those taught by teachers with bachelor's degrees. This finding was not present in the models of reading teachers, although results showed a negative, but not significant, relationship (-0.12, SE=0.32). These findings are consistent with the research on effectiveness of teacher's with master's degrees, which has shown varied effects ranging from no overall effect to a negative relationship to student achievement (Clotfelter, et al., 2010).

---

[19] This finding was also tested with a continuous variable for years of experience in both the multilevel and fixed effects models (for each subject). For math, years of experience was significant at the $p<0.001$ level for both models with a coefficient of 0.41 (SE= 0.06) for the multilevel level model and a coefficient of 0.44 (SE=0.08) for the school fixed effects model. For reading, years of experience was significant in the multilevel level model at the 0.05 level with a coefficient of 0.16 (SE=0.33), and in the school fixed effects model it was significant at the $p<0.1$ level with a coefficient of 0.12 (SE=0.07).

Table 4.3. Multilevel Baseline Model of Reading Student Achievement Gains

	Multilevel Baseline Model 1		Multilevel Model 2		Multilevel Model 3	
	β	SE	β	SE	β	SE
Intercept	6.11***	0.08	5.18***	0.23	6.38***	0.39
Student						
Year			0.35***	0.07	0.23**	0.08
Student Poverty			0.04	0.13	0.05	0.14
Minority Student			-0.34**	0.13	-0.21	0.15
Male Student			0.11	0.11	0.10	0.11
ESL Student			0.81***	0.18	0.81***	0.19
Teacher						
%Class_Minority					-0.71	0.41
%Class_Poverty					0.07*	0.44
%Class_ESL					0.25	0.62
Minority Teacher					-0.00	0.20
Male Teacher					-0.02	0.28
Master's Degree					-0.12	0.32
1st Year Teacher					-0.96***	0.28
2nd Year Teacher					-0.76**	0.26
3rd Year Teacher					-0.77***	0.24
4th Year Teacher					-0.60**	0.23
Student Variance	36.25		36.17		36.11	
Student Variance Explained			0.00		0.00	
Teacher Variance	3.75		3.59		3.57	
Teacher Variance Explained			0.04		0.05	
ICC	0.09		0.09		0.09	
Total Variance Explained			0.01		0.01	
Number of Students	11,328		11,328		11,328	
Number of Teachers	1,229		1,229		1,229	

***p<0.001 level, **p<0.01 level, *p<0.1 level

Table 4.4. Multilevel Baseline Model of Math Student Achievement Gains

	Multilevel Baseline Model 1		Multilevel Model 2		Multilevel Model 3	
	β	SE	β	SE	β	SE
Intercept	0.13*	0.08	0.86***	0.22	1.90***	0.38
Student Level						
Year			-0.28***	0.07	-0.46***	0.08
Student Poverty			-0.09	0.11	-0.09	0.11
Minority Student			-0.01	0.12	0.06	0.13
Male Student			-0.12	0.10	-0.11	0.10
ESL Student			0.73***	0.15	0.74***	0.16
Teacher Level						
%Class_Minority					-0.44	0.39
%Class_Poverty					0.35	0.40
%Class_ESL					-0.37	0.63
Minority Teacher					0.16	0.21
Male Teacher					0.10	0.25
Master's Degree					-0.86***	0.26
1st Year Teacher					-1.51***	0.28
2nd Year Teacher					-1.03***	0.27
3rd Year Teacher					-0.14	0.25
4th Year Teacher					-0.08	0.22
Student Variance	27.43		27.35		27.33	
Student Variance Explained			0.00		0.00	
Teacher Variance	5.91		5.85		5.48	
Teacher Variance Explained			0.01		0.07	
ICC	0.18		0.18		0.17	
Total Variance Explained			0.00		0.02	
Number of Students	12,605		12,605		12,605	
Number of Teachers	1,282		1,282		1,282	

***p<0.001 level, **p<0.01 level, *p<0.1 level

The last model in this section examined a baseline school fixed effects model (model 4) and a school fixed effects model with student and classroom/teacher covariates (model 5) in both subjects. Results are found in Table 4.5. The school fixed effects model examined variation between teachers within the school itself, with covariates for student, classroom, and teacher. When I examined the reading and math baseline models, the teacher-level variance was 7.79 in reading, and in math it was 10.16. In this model, all between school variance was removed by the addition of the school fixed effect, and therefore this is a measure of the within-school teacher level variance. In model 5, the teacher variance was 7.60 for reading and 10.27 in math. Given these findings between models 4 and 5, the teacher variance explained was small in both subject (at 0.02 in reading and -0.01 in math). Furthermore, the total variance explained was 0.01 for model 5 (in both subjects), suggesting the variance was not well explained by the covariates.

Findings in model 5 show that students with limited English proficiency had almost a full point increase in gains in both subjects compared to students that are proficient in English (0.80 SE=0.21 in reading and 0.74 SE=0.17 in math) (significant at the $p<0.001$ level). Additionally, in the reading model, there was a negative relationship between student gain scores and minority students, with a coefficient of -0.23 (SE=0.14), though only at the $p<0.1$ level. Results in Table 4.5 below show that early career teachers had a significant negative relationship in both subjects. A teacher's first year had a large negative coefficient of -0.67 (SE=0.33) for reading (at the $p<0.05$ level) and was -1.55 (SE=0.30) for math at the $p<0.001$ level. The coefficient declined in the second year for both subjects and dropped in significance in reading to $p<0.1$ (-0.50 SE=0.29), though remained significant at the $p<0.001$ level in math (-1.09 SE=0.29). While relationships consistently became less negative in year 3, results were only significant for reading at the $p<0.05$ level with a coefficient of -0.47(SE=0.22), and results were not significant for those with 4 years of experience in either subject.

Table 4.5. School Fixed Effects Baseline Models of Student Achievement Gains (By Subject)

	School Fixed Effects Model 4 (Reading)		School Fixed Effects Model 5 (Reading)		School Fixed Effects Model 4 (Math)		School Fixed Effects Model 5 (Math)	
	β	SE	β	SE	β	SE	β	SE
Intercept	6.05***	0.00	5.79***	0.48	0.15***	0.00	1.92***	0.57
Student Level								
Year			0.22**	0.09			-0.50***	0.10
Student Poverty			0.06	0.12			-0.13	0.12
Minority Student			-0.23*	0.14			0.02	0.15
Male Student			0.11	0.11			-0.11	0.10
ESL Student			0.80***	0.21			0.74***	0.17
Teacher Level								
%Class_Minority			0.04	0.62			-0.45	0.67
%Class_Poverty			-0.16	0.62			0.64	0.63
%Class_ESL			0.22	0.69			-0.40	0.78
Minority Teacher			0.23	0.27			0.04	0.39
Male Teacher			-0.02	0.40			-0.17	0.25
Master's Degree			-0.51*	0.31			-0.65**	0.31
1st Year Teacher			-0.67**	0.33			-1.55***	0.30
2nd Year Teacher			-0.50*	0.29			-1.09***	0.29
3rd Year Teacher			-0.47**	0.22			-0.05	0.33
4th Year Teacher			-0.29	0.24			0.00	0.25
Student Variance	39.06		38.91		31.91		31.41	
Student Variance Explained			0.00				0.02	
Teacher Variance	7.79		7.60		10.16		10.27	
Teacher Variance Explained			0.02				-0.01	
ICC	0.17		0.16		0.24		0.25	
Total Variance Explained			0.01				0.01	
Number of Students	11,328				12,605			
Number of Teachers	1,229				1,282			

***p<0.001 level, **p<0.01 level, *p<0.1 level

4.2.3 University Models

The data on the baseline models provided an initial understanding of covariates within the study and the relationships between the student, classroom, and teacher variables before the topics of interest were further investigated. The following results builds off that analysis and investigates the relationship between the university where the teacher was trained and the student achievement gains of his/her students in reading and/or math. Findings from this analysis also identified which institutions to survey for the analysis of training elements of quality and student achievement gains. These results are reported below in Table 4.6 (reading) and Table 4.7 (math).

The dataset used in this study included 41 universities that trained elementary teachers in North Carolina. However as many of those universities had less than 50 teachers, I only included institutions that initially had at least 100 teachers[20]. This reduced the number of institutions to 30 for both reading and math. For math and reading, the analysis included the same 29 schools, however 2 additional schools were added (1 to reading and 1 to math). I examined two different models for each subject. The first, model 6, was a multilevel model which added the universities where teachers received their training to model 3 and student covariates as well as teacher classroom covariates. In model 7, I examined a school fixed effects model, with university and all student and teacher classroom covariates.

[20] I identified these programs before identifying which teachers taught tested and non-tested subjects. Once I identified those teachers and programs, there was a natural reduction in the number of teachers from each institution.

As shown in Table 4.6 and Table 4.7, the variance of model 6 was not substantially different from the original baseline models. For reading gain score models with university covariates, the student variance was 36.14 and teacher classroom variance moved marginally to 3.26 (shown in Table 4.6). One improvement however in this model from the baseline model was that 13 percent of the teacher-classroom variance was explained by the addition of university covariates (while 0 percent of level-1 variance was explained by added covariates). The math gain score model with university covariates, has a similar pattern as the reading model where the student variance was reduced to 27.31, while the teacher-classroom variance was slightly reduced from the baseline model to 5.42. Additionally, 8 percent of the teacher-classroom variance was explained by the university covariates (though these covariates had no change in level-1 variance) (shown at the bottom of Table 4.7).

In the fixed models, the teacher- variance was substantially smaller. For instance, in the school fixed effects model, model 7, the teacher variance was 7.63 in reading, wherein 2 percent of the teacher variance was explained by the presence of the covariates. In model 7 for math, the variance changed to 10.47, though none of the teacher variance was explained by the covariates. Furthermore, the total variance explained for the multilevel model, model 6 in both reading and math was 0.02. For the school fixed effects model, model 7, the total variance explained was 0.00 for reading and math. Overall, these findings suggest university covariates explained some of the school variance in student gain scores compared to the baseline model but only a small percentage of total variance.

Turning to model covariates, results showed that those that were significant in models 1-3 were again significant: year, English language learners, and teacher’s year of experience. Additionally, findings showed that gains in math scores decreased by almost a full point for those taught by a teacher with a master's degree and was significant at the $p<0.001$ level with a coefficient -0.93 SE=0.27. Another difference was that the class percentage of minority students had a negative relationship to student gains in reading with a coefficient of -0.90 SE=0.43 (significant at the $p<0.05$ level) (this result was only present in model 6).

The biggest variation in the analysis of universities was between subjects, as several universities showed significant findings for those with reading gain scores, but no university had a significant relationship to math gain scores in the multilevel model. However, school fixed effects models showed significant findings for both subjects (though at a lower significance level in math). For a program to be included in the survey, the program had to have a significant effect (at the $p<0.1$ level or higher) in either both of these analyses or in one analysis per subject (thus significant in one model for reading and one model for math). I identified 20 institutions that were significant in both the multilevel and the fixed effects model in reading. University 13, University 18, and University 14 were significant at the $p<0.001$ level (in model 6) and had large coefficients. For instance, University 13 had a coefficient of 3.30 (SE=0.91) in model 6 and 4.26 (SE=1.28) in model 7. The coefficients for University 18 was 3.13 (SE=0.91) in model 6 and then 3.54 (SE=1.26) in model 7, and for University 14 the coefficients were 2.94 (SE=0.96) in model 6 and 3.32 (SE=1.33) in model 7. Additionally, University 21and University 26 were significant in the school fixed effects model (model 7) for both math and reading. There were two institutions in math that were significant in the school fixed effects model but were not significant in reading (institutions 17 and19). These were only significant in one model, model 7 and were thus not included in the subsequent analysis. Table 4.6 and Table 4.7 below have more information on these models and the coefficients of each institution included in the analysis. Subsequent analyses in this chapter only include the 24 institutions identified in this section.[21]

Table 4.6. Reading Achievement Gains by Training University[22]

	Multilevel Model Model 6		School Fixed Effects Model Model 7	
	β	SE	β	SE
Intercept	4.85***	0.85	3.69***	0.86
Student Level				
Year	0.22**	0.08	0.22**	0.10
Student Poverty	0.06	0.14	0.07	0.13
Minority Student	-0.21	0.15	-0.23*	0.14
Male Student	0.11	0.11	0.12	0.11
ESL Student	0.82***	0.19	0.81***	0.21
Teacher Level				
%Class_Minority	-0.90**	0.43	-0.31	0.63
%Class_Poverty	0.03	0.46	-0.04	0.63
%Class_ESL	0.53	0.63	0.72	0.63
Minority Teacher	-0.19	0.24	-0.00	0.29
Male Teacher	0.01	0.28	-0.02	0.38
Master's Degree	-0.08	0.32	-0.47*	0.28
1st Year Teacher	-0.95***	0.28	-0.71**	0.34
2nd Year Teacher	-0.73**	0.26	-0.49*	0.31
3rd Year Teacher	-0.77**	0.25	-0.48*	0.26
4th Year Teacher	-0.59**	0.23	-0.29	0.25
University 13	3.30***	0.91	4.04**	1.29
University 18	3.13***	0.91	3.26**	1.02
University 14	2.94***	0.96	3.41**	1.17
University 30a	2.76**	0.97	3.99***	1.09
University 25	2.59**	1.13	3.56**	1.30
University 20	2.34**	1.02	3.26**	1.56
University 23	2.34**	0.94	3.04**	1.50

[21] During the preliminary analysis of results, 2 institutions initially showed statistical significance and thus received the survey and subsequently responded to the survey (University 22 and University 28). However, in the final analysis these institutions were no longer significant, but they responded to the survey and therefore their responses are included in the analysis, for a total number of 24 institutions.

[22] Universities are sorted by the coefficient in Table 4.6 model 6.

Table 4.6 Continued

	Multilevel Model Model 6		School Fixed Effects Model Model 7	
	β	SE	β	SE
University 24	2.19**	1.03	2.96***	0.85
University 2	1.81**	0.77	2.31**	0.93
University 15	1.74*	0.99	1.69*	1.69
University 9	1.66**	0.84	2.20**	0.94
University 3	1.63**	0.77	2.20**	0.94
University 1	1.61**	0.77	2.24**	0.89
University 28	1.61	1.34	2.79	2.04
University 22	1.57	1.09	2.24**	1.03
University 29	1.57*	0.91	2.93**	0.96
University 5	1.51*	0.79	1.84**	0.93
University 8	1.47*	0.84	1.93*	1.56
University 27	1.46*	0.95	2.24**	1.04
University 6	1.34*	0.84	2.12**	1.02
University 10	1.34*	0.83	2.48**	1.02
University 21	1.32	1.00	2.02*	1.20
University 11	1.27	1.05	1.40	1.09
University 16	1.24	0.94	2.46**	1.25
University 4	1.22*	0.77	1.77**	0.76
University 17	1.22	1.00	1.89	1.27
University 19	1.03	1.05	1.03	1.24
University 26	1.01	0.95	2.16**	1.09
University 12	0.82	0.90	1.14	1.11
University 7	0.64	0.88	1.26	0.99
Student Variance	36.14		38.76	
Student Variance Explained	0.00		-0.01	
Teacher Variance	3.26		7.63	
Teacher Variance Explained	0.13		0.02	
ICC	0.08		0.16	
Total Variance Explained	0.02		0.00	
Number of Students	11,328		11,328	
Number of Teachers	1,229		1,229	

***p<0.001 level, **p<0.01 level, *p<0.1 level

Table 4.7. Math Achievement Gains by Training University[23]

	Multilevel Model Model 6		School Fixed Effects Model Model 7	
	β	SE	β	SE
Intercept	1.02	1.07	0.47	0.82
Student Level				
Year	-0.44***	0.08	-0.50***	0.09
Student Poverty	-0.08	0.11	-0.12	0.12
Minority Student	0.06	0.13	0.03	0.15
Male Student	-0.12	0.10	-0.12	0.10
ESL Student	0.74***	0.16	0.75***	0.17
Teacher Level				
%Class_Minority	-0.43	0.42	-0.31	0.71
%Class_Poverty	0.38	0.43	0.50	0.67
%Class_ESL	-0.49	0.66	-0.44	0.83
Minority Teacher	0.15	0.25	0.04	0.41
Male Teacher	0.10	0.25	-0.20	0.25
Master's Degree	-0.93***	0.27	-0.75**	0.31
1st Year Teacher	-1.51***	0.28	-1.62***	0.29
2nd Year Teacher	-1.04**	0.27	-1.17***	0.29
3rd Year Teacher	-0.16	0.25	-0.15	0.32
4th Year Teacher	-0.11	0.23	-0.09	0.24
University 21	1.83	1.35	2.35*	1.39
University 24	1.52	1.27	2.21**	1.07
University 13	1.52	1.14	2.46*	1.31
University 10	1.34	1.12	2.16**	1.01
University 17	1.30	1.22	1.95*	1.20
University 23	1.20	1.34	1.16	1.21
University 19	1.16	1.21	1.70*	1.11
University 15	1.10	1.17	1.63*	0.87
University 6	1.03	1.09	1.89**	0.90
University 2	1.00	1.03	2.05**	0.80
University 26	0.90	1.23	2.54*	1.41
University 25	0.90	1.19	1.26	1.10
University 9	0.90	1.07	1.70**	0.80
University 1	0.89	1.03	1.32*	0.74
University 20	0.88	1.30	2.48*	1.66
University 11	0.87	1.14	0.93	0.93
University 18	0.87	1.13	1.50*	0.90
University 16	0.86	1.21	0.47	0.91
University 4	0.85	1.04	1.28*	0.82
University 8	0.76	1.09	1.74*	0.93

[23] Universities are sorted by the coefficient in Table 4.7 model 6.

Table 4.7 Continued

	Multilevel Model Model 6		School Fixed Effects Model Model 7	
	β	SE	β	SE
University 22	0.74	1.14	-0.36	0.90
University 27	0.72	1.17	1.67*	0.97
University 3	0.72	1.04	1.15	0.80
University 12	0.63	1.11	1.19	1.00
University 29	0.63	1.17	1.35	0.94
University 7	0.52	1.15	0.76	1.07
University 30b	0.35	1.16	0.93	0.79
University 5	0.35	1.05	1.53*	0.86
University 14	0.19	1.16	1.50	1.07
University 28	-1.13	1.35	-0.12	0.93
Student Variance	27.31		31.40	
Student Variance Explained	0.00		0.02	
Teacher Variance	5.42		10.47	
Teacher Variance Explained	0.08		-0.00	
ICC	0.17		0.25	
Total Variance Explained	0.02		0.00	
Number of Students	12,605		12,605	
Number of Teachers	1,282		1,282	

***p<0.001 level, **p<0.01 level, *p<0.1

4.2.4 Full Analysis of Program Elements

Question two of this study examines whether program elements had an impact on student gain scores for either reading or math. Once I identified the above institutions that showed a significant relationship to student achievement gains, I tested the relationship between student gains in reading and math using the variables gathered from the document review analysis as well as the survey (identified earlier in Table 4.1 and Table 4.2). For this analysis, I treated variables from both the document review and survey equally, however I used three statistical processes to determine which elements to add into the final model. First, I added each element into the multilevel model individually (with student, classroom and teacher covariates) to see if the element itself had any effect on student gains. For some variables, I also tested a continuous variable for the element (such as number of courses) as well as a dichotomous variable (such as high/low number of courses as determined by the top and bottom quartile)[24]. I then added elements into the model in blocks based on topic: faculty, candidate teaching experiences, required courses, assessment of candidate teaching, technology, mentoring, and noted program attributes (combined with student and teacher-classroom covariates)[25]. These initial findings are presented in Table 4.8 for both subjects.

As I entered variables that showed a significant relationship from the individual and blocked analysis into the full model, several variables were collinear. Therefore, I examined a correlation plot to test relationships between program elements and make a final determination on which variables to add to the final models. The correlation plot of program elements is shown in Figure 4.1. Results for full multilevel models (models 8 and 9) are shown in Table 4.9 and Table

---

[24] For this analysis, high and low numbers of courses were made using the top and bottom quartile for each variable. For Technology courses, high was 10 or more courses, and low was 3 or fewer. For courses that record the candidate's teaching practice, high was 6 or more courses and low was 2 or fewer. Number of courses with active teaching, high number of courses was defined as 15 or more courses, and low was defined as 5 or fewer

[25] Although only 16 of the 24 programs responded to the survey, I collected data on program requirements from curriculums from all 24 schools. As such, I ran the individual analyses for each program variable with those 24 schools. However, when variables were combined (using both those from program curriculums and those from survey responses), I only used the 16 programs that responded.

4.10, and results for the school fixed effects models (models 10 and 11) with all elements and covariates are presented in Table 4.11 (reading) and Table 4.12 (math)[26].

4.2.5 Reading

Results from the first analysis where elements were entered into the model individually and in blocks by topic with student and teacher-classroom covariates, varied substantially between subjects. For instance, there were several elements that were significant in the reading model yet only two elements were significant above the $p<0.1$ level in the models of math gains (and this was only when entered into the model in blocks). The first significant element in the reading model was the percentage of adjunct faculty, which was significant at the $p<0.1$ level with a coefficient of 0.29 (SE=0.18) and became more significant at the $p<0.05$ level when entered into the model with other faculty specific variables (Table 4.8). Additionally, while no other faculty element was significant when entered individually, two additional faculty variables were significant when entered into the model in blocks: programs where adjuncts teach all course types and programs that have multiple incentives to improve faculty teaching. Although, variable for type of course taught by adjunct was only significant at the $p<0.1$ level, faculty incentives had a significance at the $p<0.05$ level with a coefficient of 0.81 (SE=0.40).

Only one variable was significant for reading in block 2 (candidate teaching opportunities), low number of courses with active teaching opportunities. When entered individually in the model,

[26] Additionally, although data on overall programmatic changes and specific changes in elements that occurred during the survey period were collected (such as adoption of edTPA) when these were tested in the model, no notable changes were seen and therefore those results are not included.

this was significant at the $p<0.05$ level at 0.83 (SE=0.37), however it became less significant ($p<0.1$) when entered into the model with other variables on candidate teaching experiences. In block 3, required courses, the first variable that saw significance pertained to the average SAT scores of the program. For instance, while the average SAT score was significant the $p<0.05$ level, it has a coefficient of -0.00, as such I tested a dichotomous variable for competitive/high SAT and not-competitive/low SAT institutions. Results showed that not-competitive institutions had a positive coefficient of 0.80 (SE=26) at the $p<0.05$ when entered individually and was also significant when entered in blocks at the $p<0.1$ level. The three course types that were significant were: number of method courses, English courses, and foundation courses. The number of method courses was slightly significant negative relationship to gain scores when entered individually and gained in significance when entered in blocks at the $p<0.05$ level with a coefficient of -0.66 (SE=0.23). The number of English courses was significant when entered individually (with a coefficient of 0.21 with a SE=0.09 at the $p<0.05$ level), though became less significant at the $p<0.1$ level when entered with blocks of program requirements. Additionally, the number of foundations in education courses changed in significance as it was significant at the $p<0.05$ level when entered into the model individually and though was only significant at the $p<0.1$ level when entered into the model in blocks. The coefficient also slightly increased between iterations from 0.30 (SE=0.13) to 0.38 (SE=0.21).

Only one element saw a significant relationship in block 4: number of technology courses. However, this was only significant at the $p<0.1$ level (0.31, SE=0.19) and only when entered in as blocks. For block 5, assessment of candidate teaching, two variables were significant. The first was low numbers of courses that videotape teaching practice. While the number of courses that videotape a candidate's teaching wasn't' significant in any iteration, low number of courses (2 or less) was significant both when entered individually and when entered with other teaching assessment variables at the $p<0.05$ level (and increased in the size of the relationship from 0.41 SE=0.20 in the individual model to 0.68 SE=0.27 in the blocked model). Finally, assessment using pre and post-tests was significant at the $p<0.001$ level. When entered individually, the relationship to gain scores was 0.95 (SE=0.25) and increased to 1.02 (SE=0.26) when entered into the model in blocks. These findings suggest that the use of pre and post-tests in candidate teaching reviews have an increase in achievement gains by one point in reading.

Finally, variables in block 7 were designed to test what elements students, faculty, and/or program coordinators saw as the most valuable attribute of the program's training. This was split into 4 variables: student teaching experience, student diversity, faculty relationships, and preparation for teaching exams. When entered individually student teaching experience had a large negative significant relationship at -1.24 (SE=0.31) at the $p<0.001$ level and decreased slightly when entered with other attributes to -0.91 (SE=0.36) at the $p<0.05$ level (Table 4.8). Similarly, student diversity had a strong positive relationship to student reading gains when entered individually at 0.77 (SE=0.23) $p<0.001$ and decreased in size and significance when entered as blocks to 0.59 (SE=0.24) $p<0.05$.

4.2.6 Math

As noted previously, fewer elements are significant in math than in reading, and those that are have a significance at the $p<0.05$ level or higher. There were no elements within the first block on faculty that showed a significant relationship to student gain scores. Additionally, in block 2, elements that pertained to candidate teaching experiences, only the variable for low number of courses with active teaching had a significant relationship. This element was only significant when entered individually in the model with other student and teacher-classroom covariates at the $p<0.1$ level (0.61 SE=0.38).

There were more elements that were significant in block 3, required courses, than in other groupings for math gain scores. The first element to show a significant relationship was that of the number of courses on assessment of a student's work. This was not significant when entered individually but was when entered with other variables in this block at the $p<0.05$ level with a coefficient of 0.70 (SE=0.28). The next element to show significance was the number of math only courses, though was also only significant when entered into the model in blocks with a negative relationship of -0.65 (SE=0.34), at the $p<0.1$ level. Finally, similarly to reading gain scores, the number of foundations in education courses saw a positive relationship in the blocked models, though at the $p<0.1$ level (0.43, SE=0.23) (Table 4.8).

The last two blocks with significant relationship between program elements and math gain scores was technology and techniques used to assess candidate teaching. For instance, in block 4 the number of courses that incorporated instructional technology, was significant both when entered individually and in blocks at the $p<0.1$ level. Unlike other variables, the size of the relationship barely changed between iterations as the coefficient was 0.28 (SE=0.15) when entered individually and 0.29 (SE=0.15) in blocked models. I further tested this element with a dichotomous variable to test whether a higher or lower number of courses made a difference and the variable for low number of courses with instructional technology was significant. When entered individually, low number of courses had a negative significant relationship of -0.38 (SE=0.21) at the $p<0.1$ level and a larger negative effect when entered in blocks with a coefficient of -0.51 (SE=0.25) at the $p<0.05$ level. Finally, for the assessment of a candidate's teaching, programs with a formal teaching assessment had a significant effect in both the individual and blocked models, at the $p<0.1$ level (0.36 SE=0.23 in the individual model increasing to 0.43 SE=0.27 in the blocked model.

Table 4.8. Exploratory Models of Achievement Gains by Program Elements and Subject

	Elements Assessed Individually (Reading)		Elements Assessed in Blocks (Reading)		Elements Assessed Individually (Math)		Elements Assessed in Blocks (Math)	
	β	SE	β	SE	β	SE	β	SE
Block 1: Faculty								
Number FTE Faculty	-0.00	0.00	-0.00	0.01	0.00	0.00	-0.00	0.01
Number Adjunct	-0.00	0.00	-0.00	0.00	0.00	0.00	0.00	0.00
Percentage of Adjunct	0.29*	0.18	0.57**	0.24	-0.22	0.20	-0.17	0.27
Adjunct Teach all Course Types	0.01	0.02	0.05*	0.03	0.01	0.02	0.04	0.03
Faculty Incentives for Teaching	-0.01	0.21	0.81**	0.40	0.23	0.22	0.49	0.39
Block 2 Candidate Teaching								
Early Candidate Teaching Experience	-0.17	0.17	0.18	0.39	-0.13	0.18	0.23	0.44
Classroom Experience in Intro Course	-0.09	0.26	-0.04	0.37	0.29	0.26	0.53	0.39
Number of Active Teaching	-0.04	0.18	0.11	0.20	-0.17	0.20	-0.11	0.23
High Number of Active Teaching	0.26	0.22	-0.03*	0.23	-0.04	0.256	0.03	0.27
Low Number of Active Teaching	0.83**	0.37	0.76*	0.40	0.61*	0.38	0.55	0.45
Computer Adaptive Teaching	-0.18	0.20	-0.28	0.33	-0.00	0.22	-0.35	0.37
Hours of Student Teaching	-0.00	0.00	-0.00	0.00	-0.00	0.00	-0.00	0.00
Block 3: Required Courses								
Average SAT Score (University)	-0.00**	0.00	-0.00	0.00	-0.00	0.00	0.00	0.00
High SAT University	-0.07	0.17	0.57	0.43	-0.10	0.17	-0.28	0.48
Low SAT University	0.80**	0.26	0.66*	0.35	0.02	0.31	-0.54	0.49
Hours Prior to Student Teaching	-0.00	0.00	0.00	0.00	0.00	0.00	-0.00	0.00
Method Courses	-0.14*	0.10	-0.66**	0.23	-0.02	0.10	0.06	0.25
Assessment Courses	0.10	0.10	-0.05	0.24	0.09	0.11	0.70**	0.28
Math Only Courses	0.06	0.12	0.02	0.31	-0.00	0.13	-0.65*	0.34
Teaching Math Courses	-0.01	0.13	0.93*	0.49	-0.02	0.12	-0.61	0.61
English Only Courses	0.21**	0.09	0.37*	0.24	0.04	0.10	0.11	0.27
Teaching Reading Courses	0.17*	0.10	0.01	0.17	0.02	0.11	-0.23	0.18

Table 4.8 Continued

	Elements Assessed Individually (Reading)		Elements Assessed in Blocks (Reading)		Elements Assessed Individually (Math)		Elements Assessed in Blocks (Math)	
	β	SE	β	SE	β	SE	β	SE
Block 3: Required Courses (Cont.)								
Diverse Learning Courses	-0.01	0.09	-0.27	0.21	0.03	0.10	0.06	0.24
Foundation Courses	0.30**	0.13	0.38*	0.21	0.15	0.13	0.43*	0.23
Block 4: Technology								
Number of Courses with Technology	0.07	0.14	0.32	0.14	0.28*	0.15	0.29*	0.15
High Number of Courses with Tech	0.12	0.26	0.34	0.29	0.32	0.29	-0.02	0.33
Low Number of Courses with Tech	-0.08	0.20	0.20	0.25	-0.38*	0.21	-0.50**	0.25
Number of Technology Courses	0.18	0.13	0.31*	0.19	0.02	0.13	-0.16	0.22
Block 5: Assessment of Candidates								
Number of Courses that Record Teaching	-0.04	0.04	-0.00	0.04	0.03	0.05	0.02	0.05
High Number of Courses Record	0.16	0.20	0.74**	0.27	0.11	0.22	0.28	0.30
Low Number of Courses Record	0.41**	0.20	0.68**	0.27	0.08	0.23	0.20	0.31
Formal Teaching Assessment	-0.05	0.22	-0.26	0.25	0.36*	0.23	0.43*	0.27
Course Based Assessment of Candidates	0.02	0.19	-0.11	0.21	0.25	0.21	0.37*	0.22
Pre-Post Student Tests	0.95***	0.25	1.02***	0.26	0.29	0.30	0.16	0.32
Block 6: Mentoring								
Mentor in In-Service Training	-0.26	0.21	-0.31	0.28	-0.04	0.22	-0.01	0.29
Mentor Review of Teaching Practice	-0.13	0.19	0.07	0.26	-0.06	0.21	-0.05	0.27
Block 7: Noted Program Attributes								
Student Teaching Experience	-1.24***	0.31	-0.91**	0.36	-0.04	0.33	-0.15	0.38
Diversity	0.77***	0.23	0.59**	0.24	-0.13	0.23	-0.16	0.25
Faculty	0.42	0.36	-0.08	0.39	-0.08	0.41	-0.09	0.45
Preparedness for State Exams	0.33*	0.22	0.28	0.22	-0.00	0.26	-0.04	0.27

***p<0.001 level, **p<0.05 level, p<0.1 level

After identifying elements that had a significant relationship in reading and/or math samples, I began to enter these into the full models, however, several elements showed a collinear relationship with one another. As a result, I examined a correlation plot to understand the relationship between these elements. From that review, I was then able to determine which elements should be added to the full model. A correlation plot for elements identified as having a significant effect above in Table 4.8, is shown below in Figure 4.1.

In the initial analysis of the correlation plot, I found that many of these elements were correlated with one another. Indeed, of the 25 elements originally tested, all have at least one variable with which it was positively or negatively correlated with a correlation of 0.50 or higher. The first block I assessed was that of faculty. In this block all 3 elements were strongly correlated with 3 other elements. After reviewing this, I assessed which elements were correlated with one another and determined which should remain in the full model. For instance, adjunct faculty variables were correlated with the elements of technology, formal assessment, teaching reading, teaching methods, intro teaching experiences, and math courses. Each of these elements were all necessary to include given their own significance in this study and prior literature when compared to that of adjunct variables. Whereas in the case of faculty incentives for teaching, this variable is strongly correlated with teaching experiences in the intro to teaching course, however, I have other variables that measure this element but no other variables that measure/assess faculty teaching. As such, I removed all variables in this block except for faculty incentives. For the second block of elements, *candidate teaching experiences*, courses with active teaching components were correlated with assessment courses, teaching math courses, and foundation in education courses, many of which tend to have an in-classroom experience. I reviewed relationships between these elements and removed courses specific to student assessment as it was strongly correlated with other elements.

The correlation plot showed that many elements within block 3, *required courses,* were highly correlated with one another. For example, courses on teaching methods (that are not subject specific) were correlated with courses on teaching math as well as subject specific reading courses. Ultimately, I kept method courses in the full model given it is especially pertinent to the research question. However, I tested model changes with and without this element to better understand the relationship and changes were minimal. The only element that I removed from this section (other than assessment) was diverse learning courses as it was highly correlated with courses that taught math and reading methods. For courses in block 4, *technology,* I kept the number of courses that had lessons using instructional technology (instead of number of courses that focused solely on the use of instructional technology) as this was significant in the individual and blocked models and not highly correlated with other variables.

The last two blocks that I assessed were those in block 5 (*assessment of candidate teaching*) and block 7 (*noted program attributes*). As shown in the correlation plot, the program element of course based assessment of candidates (and not formal assessment outside of courses), was highly correlated with the notable attribute in block 7 unique/diverse student body. These two variables were both unique to this study, however when I entered them both into the full model, no element was significant. When I removed both elements there were variables that were significant. Furthermore, when either variable was in the model (together or separately), the model showed low power as several coefficients were very large compared to coefficients of other variables (above 5.00) yet were not significant. I therefore removed these elements in the full model. Finally, the variable for the candidate's favorite program aspect was student teaching was strongly correlated with pre and post-tests in candidate assessment. Given the significance of the relationship of the pre-post-test element, I excluded that element.

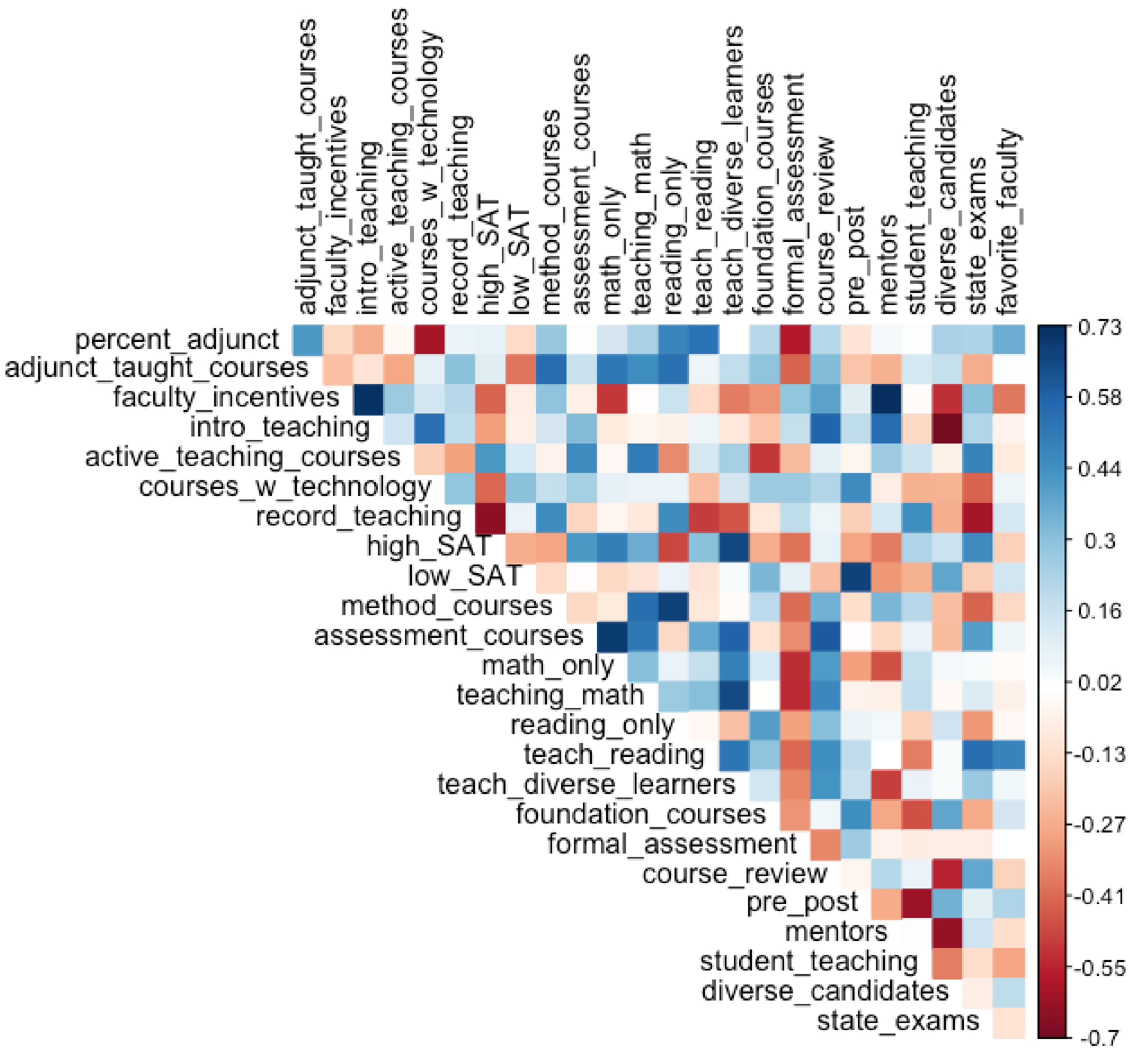

Figure 4.1 Correlation Plot of Program Elements

4.2.7 Full Models

Once I determined which variables to include in the full model, I ran three multilevel models to test the relationship between program elements and student gain scores. The first was a baseline model, model 8 (as the sample for this analysis only included those trained by institutions that participated in the survey). I then ran two full models: model 9 a multilevel model with all program elements, and model 10 a full multilevel model with dichotomous variables of high/low courses as examined in the individual and blocked models. Table 4.9 shows results for the analysis in reading and Table 4.10 shows results for the analysis in math. School fixed effects models were also examined and those are discussed in the section following the multilevel models (and results are shown in Table 4.11and Table 4.12 for reading and math).

Before examining the full models of training elements and student achievement gain scores, I assessed the change in variance for models with all selected program elements for both subjects. For model 9, the full multilevel model, the variance was not substantially different from the baseline model for either subject, however the proportion of teacher-classroom variance explained by these covariates changed from the baseline model by 17 percent for reading and 11percent in math. Additionally, the total variance explained was 0.02 in both subjects (Table 4.9 and Table 4.10). In the school fixed effects model, the proportion of teacher variance explained by these covariates was 0.00 in reading models and -0.05 in math (model 12). Furthermore, the total variance explained for the fixed models was 0.01 in reading and 0.00 in math. Results for these models are at the bottom of Table 4.11 (reading) and Table 4.12 (math).

Results from the full models showed large variations between subjects. As shown in Table 4.9 below, two program elements were significant in both models in reading (models 9 and 10). The number of courses on methods had a negative significant relationship in model 9 with a coefficient of -0.57 (SE=0.38) at the p<0.1 level and grew in significance in model 10 with dichotomous variables to -0.86 (SE=0.33) at the p<0.05 level. Thus, one standard deviation increase in method courses, would lead to a decrease in gains in a student's achievement in reading by -0.11 standard deviations. Additionally, the number of English courses was also significant in both models. In model 9, this element was significant at the p<0.05 level and had a coefficient of 0.65 (SE=0.22) and similar to method courses it increased in significance in model 10 to p<0.001 with a coefficient of 0.67 (SE=0.20) (Table 4.9). Findings suggest that one standard deviation increase in English content courses during a candidate's training increases that student achievement gains in reading by 0.09 standard deviations. Additionally, low number of courses with instructional technology[27] was significant in model 10, This element had a slight negative relationship at -0.84 (SE=0.57) at the p<0.1 level.

In the analysis of math gain scores, the program element with the highest significance was the number of foundation courses (Table 4.10). This element was significant at the p<0.05 level showed almost a whole point increase in math gain scores at 0.77 (SE=0.35). Furthermore, with every increase of one standard deviation in foundation courses in a teacher's training, student achievement gains increased by 0.11 standard deviations. However, this was not present in model 10 with dichotomous variables. One variable that was significant in model 10 that had not shown significance previously was faculty incentives. This element was significant at the p<0.05 level with a coefficient of 0.64 (SE=0.33). Thus, one standard deviation increase in faculty incentives,

[27] For models10 and 13, I tested whether a low number of courses in 3 areas (courses with teaching experiences, courses with instructional technology imbedded within, and courses that recorded candidate teaching) had a relationship with student scores. High number of courses were tested as well, however this saw no significant findings and as such only results for low number of these courses are shown.

leads to an increase in math student achievement gains by 0.05 standard deviations. Results for these models in math are shown below in Table 4.10.

Table 4.9. Multilevel Models of Reading Achievement Gains by Program Elements

	Baseline Model 8		Multilevel Model 9		Multilevel Model 10	
	β	SE	β	SE	β	SE
Intercept	6.10***	0.09	5.07***	1.50	7.18***	1.19
Student						
Year			0.26**	0.09	0.30**	0.09
Student Poverty			-0.03	0.17	-0.02	0.17
Minority Student			-0.09	0.20	-0.09	0.20
Male Student			0.09	0.14	0.09	0.14
ESL Student			0.69**	0.25	0.69**	0.25
Teacher						
%Class_Minority			-0.66	0.53	-0.67	0.53
%Class_Poverty			-0.77	0.56	-0.70	0.56
%Class_ESL			1.36*	0.76	1.33*	0.76
Minority Teacher			-0.54**	0.26	-0.51**	0.26
Male Teacher			-0.49	0.34	-0.50*	0.34
Master's Degree			-0.20	0.38	-0.18	0.38
1st Year Teacher			-0.77**	0.34	-0.78**	0.34
2nd Year Teacher			-0.78**	0.34	-0.78**	0.33
3rd Year Teacher			-0.63**	0.30	-0.62**	0.30
4th Year Teacher			-0.60**	0.30	-0.60**	0.28
Faculty Incentives			-0.07	0.38	0.08	0.30
No. Active Teaching			0.65	0.47		
Low No. Active Teaching					-0.26	0.81
Low SAT University			0.04	0.49	0.63	0.57
Method Courses			-0.57*	0.38	-0.86**	0.33
Math Only Courses			-0.09	0.31	-0.26	0.27
Teaching Math Courses			-0.23	0.55	0.29	0,.40
English Only Courses			0.65**	0.22	0.67***	0.20
Teaching Reading Courses			-0.05	0.21	-0.15	0.27
Foundation Courses			0.30	0.27	0.05	0.30
No. Courses with Technology			0.02	0.30		
Low No. Courses with Tech					-0.84*	0.57
No. Courses that Record			0.02	0.07		
Low No. Courses Record					0.63	0.55
Formal Teaching Assessment			-0.21	0.66	-0.72	0.53
Pre-Post Student Tests			0.47	0.62	-0.17	0.64
Student Variance	36.00		35.88		35.88	
Student Variance Explained			0.00		0.00	
Teacher Variance	3.22		2.67		2.66	
Teacher Variance Explained			0.17		0.17	
ICC	0.08		0.07		0.07	
Total Variance Explained			0.02		0.02	
Number of Students	7,191		7,191		7,191	
Number of Teachers	741		741		741	

***p< 0.001 level. **p< 0.05 level * p<0.1 level

Table 4.10. Multilevel Models of Math Achievement Gains by Program Elements

	Baseline Model 8		Multilevel Model 9		Multilevel Model 10	
	β	SE	β	SE	β	SE
Intercept	0.09	0.10	-1.57	2.58	-1.31	1.89
Student						
Year			-0.49***	0.10	-0.49***	0.10
Student Poverty			-0.14	0.15	-0.14	0.15
Minority Student			-0.01	0.17	-0.01	0.17
Male Student			-0.13	0.13	-0.13	0.13
ESL Student			0.80***	0.21	0.80***	0.21
Teacher						
%Class_Minority			-0.00	0.55	-0.01	0.56
%Class_Poverty			-0.14	0.57	-0.13	0.57
%Class_ESL			0.52	0.87	0.52	0.87
Minority Teacher			0.04	0.28	0.05	0.28
Male Teacher			-0.19	0.33	-0.19	0.33
Master's Degree			-1.04**	0.35	-1.05**	0.36
1st Year Teacher			-1.52***	0.37	-1.51***	0.37
2nd Year Teacher			-0.97**	0.34	-0.97**	0.34
3rd Year Teacher			-0.02	0.33	-0.02	0.33
4th Year Teacher			0.06	0.30	0.06	0.30
Faculty Incentives			0.54	0.47	0.64**	0.33
No. Active Teaching			0.21	0.90		
Low No. Active Teaching					-0.41	2.31
Low SAT University			0.21	0.62	0.11	0.74
Method Courses			0.10	0.52	-0.03	0.38
Math Only Courses			0.30	0.42	0.16	0.29
Teaching Math Courses			0.24	0.86	0.33	0.53
English Only Courses			-0.19	0.24	-0.13	0.27
Teaching Reading Courses			0.15	0.20	0.25	0.45
Foundation Courses			0.77**	0.35	0.83	0.65
No. Courses with Technology			-0.16	0.30		
Low No. Courses with Tech					0.30	0.57
No. Courses that Record			0.04	0.08		
Low No. Courses Record					-0.30	0.56
Formal Teaching Assessment			1.22	0.97	1.13	0.81
Pre-Post Student Tests			-0.58	0.64	-0.70	0.65
Student Variance	27.70		27.58		27.57	
Student Variance Explained			0.00		0.00	
Teacher Variance	5.86		5.20		5.22	
Teacher Variance Explained			0.11		0.11	
ICC	0.17		0.16		0.16	
Total Variance Explained			0.02		0.02	
Number of Students	7,724		7,724		7,724	
Number of Teachers	748		748		748	

***p< 0.001 level. **p< 0.05 level * p<0.1 level

The results for the school fixed effects models in reading showed no significant elements in model 12 (shown below in Table 4.11). However, when dichotomous elements were entered into model 13, the number of courses on methods and the number of English only courses were then significant. Similar to the multilevel model, the number of teaching method courses had a negative significant relationship at -0.66 (SE=0.40) at the $p<0.1$ level. Additionally, the number of English courses was significant at the $p<0.05$ level with a coefficient of 0.50 (SE=0.23).

The results from the math school fixed effects models showed more significant elements than in reading (results for math models are shown below in Table 4.12). For instance, the number of foundation courses was significant in both model 12 and 13, though decreased in significance between models. In model 12, foundation courses were significant at the $p<0.05$ level with a coefficient of 0.94 (SE=0.43), and the coefficient then increased in size in model 13 to 1.66 (SE=0.91), though was only significant at the $p<0.1$ level. The number of courses with instructional technology had a negative relationship to math student gain scores in model 12 with a coefficient of -0.91 (SE=0.50) at the $p<0.1$ level. On further examination, a low number of courses with instructional technology had a large positive relationship at the $p<0.05$ level and a coefficient of 1.47 (SE=0.75) (Table 4.12 model 13). The final element that showed a strong relationship to math gain scores was formal assessment of a candidate's teaching. In model 12 this relationship was 1.91 (SE=1.25) at the $p<0.1$ level and became stronger when dichotomous variables were added in model 13, increasing to 2.50 (SE=0.1.23) at the $p<0.05$. Findings showed that one standard deviation increase in formal assessment of candidate teaching translated to an increase in gains of student's math achievement by 0.12 standard deviations.

Table 4.11. School Fixed Effects of Reading Achievement Gains by Program Elements

	Baseline School Fixed Effects Model 11		School Fixed Effects Model 12		School Fixed Effects Model 13	
	β	SE	β	SE	β	SE
Intercept	6.04***	0.00	5.55**	1.86	5.94***	1.51
Student						
Year			0.21*	0.12	0.22*	0.12
Student Poverty			-0.01	0.18	-0.02	0.18
Minority Student			-0.16	0.18	-0.16	0.18
Male Student			0.06	0.15	0.06	0.15
ESL Student			0.67**	0.25	0.67**	0.25
Teacher						
%Class_Minority			0.11	0.76	0.06	0.77
%Class_Poverty			-1.33*	0.79	-1.29*	0.79
%Class_ESL			1.25	0.97	1.21	0.97
Minority Teacher			-0.28	0.36	-0.27	0.35
Male Teacher			-0.48	0.49	-0.47	0.49
Master's Degree			-0.70*	0.38	-0.72*	0.39
1st Year Teacher			-0.44	0.44	-0.42	0.44
2nd Year Teacher			-0.49	0.39	-0.47	0.39
3rd Year Teacher			-0.39	0.34	-0.36	0.33
4th Year Teacher			-0.41	0.34	-0.39	0.33
Faculty Incentives			0.26	0.58	0.14	0.46
No. Active Teaching			-0.09	0.63		
Low No. Active Teaching					0.47	1.19
Low SAT University			0.01	0.76	0.27	0.90
Method Courses			-0.77	0.54	-0.66*	0.40
Math Only Courses			-0.15	0.46	-0.13	0.34
Teaching Math Courses			0.60	0.83	0.60	0.60
English Only Courses			0.44	0.31	0.50**	0.23
Teaching Reading Courses			-0.05	0.25	-0.13	0.37
Foundation Courses			0.22	0.42	0.05	0.38
No. Courses with Technology			-0.00	0.49		
Low No. Courses with Tech					-0.26	0.86
No. Courses that Record			0.08	0.11		
Low No. Courses Record					-0.03	0.79
Formal Teaching Assessment			-0.32	0.89	-0.29	0.74
Pre-Post Student Tests			0.80	1.02	0.56	0.99
Student Variance	38.19		37.82		37.82	
Student Variance Explained			0.01		0.01	
Teacher Variance	10.04		9.99		10.11	
Teacher Variance Explained			0.00		-0.01	
ICC	0.21		0.21		0.21	
Total Variance Explained			0.01		0.01	
Number of Students	7,191		7,191		7,191	
Number of Teachers	741		741		741	

***p< 0.001 level. **p< 0.05 level * p<0.1 level

Table 4.12. School Fixed Effects of Math Achievement Gains by Program Elements

	Baseline Model 11		Fixed Model 12		Fixed Model 13	
	β	SE	β	SE	β	SE
Intercept	0.08***	0.00	-2.12	3.17	-4.28*	2.66
Student						
Year			-0.59***	0.13	-0.59***	0.12
Student Poverty			-0.18	0.15	-0.18	0.15
Minority Student			-0.04	0.19	-0.04	0.19
Male Student			-0.17	0.13	-0.17	0.13
ESL Student			0.77***	0.19	0.77***	0.19
Teacher						
%Class_Minority			0.45	0.78	0.40	0.76
%Class_Poverty			0.01	0.70	0.04	0.71
%Class_ESL			1.49	1.11	1.53	1.10
Minority Teacher			0.09	0.29	0.11	0.43
Male Teacher			-0.33	0.29	-0.32	0.28
Master's Degree			-0.51	0.42	-0.51	0.43
1st Year Teacher			-1.56***	0.38	-1.54***	0.38
2nd Year Teacher			-1.12**	0.36	-1.11**	0.36
3rd Year Teacher			-0.05	0.38	-0.05	0.38
4th Year Teacher			0.03	0.32	0.03	0.32
Faculty Incentives			0.13	0.58	0.34	0.37
No. Active Teaching			0.24	1.17		
Low No. Active Teaching					-3.46	3.72
Low SAT University			0.15	0.82	-0.89	1.09
Method Courses			0.35	0.76	-0.09	0.42
Math Only Courses			0.66	0.62	0.28	0.35
Teaching Math Courses			0.67	1.22	0.05	0.93
English Only Courses			-0.40	0.31	0.16	0.40
Teaching Reading Courses			-0.05	0.31	0.70	0.80
Foundation Courses			0.94**	0.43	1.66*	0.91
No. Courses with Technology			-0.91*	0.50		
Low No. Courses with Tech					1.47**	0.75
No. Courses that Record			0.12	0.11		
Low No. Courses Record					-0.71	0.70
Formal Teaching Assessment			1.91*	1.25	2.50**	1.23
Pre-Post Student Tests			0.57	1.17	-0.14	0.70
Student Variance	31.58		31.02		31.02	
Student Variance Explained			0.02		0.02	
Teacher Variance	8.64		9.12		9.06	
Teacher Variance Explained			-0.05		-0.05	
ICC	0.21		0.23		0.23	
Total Variance Explained			0.00		0.00	
Number of Students	7,724		7,724		7,724	
Number of Teachers	748		748		748	

***p< 0.001 level. **p< 0.05 level * p<0.1 level

5.0 Program Elements and Teacher Attrition

This chapter provides results to research question 3, which is an examination of teacher training elements and teacher attrition in public schools[28] in North Carolina. Specifically, the research question guiding this chapter is:

3) What is the relationship between elements of a teacher preparation program and elementary teacher attrition in the first 8 years of a teacher's career in North Carolina?

This analysis differed from that in chapter 4, as it only examined the career longevity of all elementary teachers in North Carolina public schools up to a teacher's eighth year of teaching, for those who received their teacher training from program in North Carolina. Data used for this analysis combined teacher employment records with data on teachers, the schools[29] where they taught, data on the universities where they received their training (as well as program elements included in their training). The first step to answer this question was to create a long form database where each teacher in the dataset had one record per year of teaching. I then used a discrete time survival analysis model to calculate the estimated hazard, or risk, of no longer teaching in public

---

[28] The data available for this study only included teacher records of public k-12 schools in North Carolina and as such this study is only able to determine if a teacher continued teaching in a public k-12 school in North Carolina.

[29] Because this analysis required longer time periods of teachers to measure career longevity, student gain scores are not included as it reduced the number of teachers beyond which the analysis could be performed. See chapter 3 for more information on the methodological approach for this analysis.

schools in North Carolina (measured by the proportion of those that dropped out of the dataset each year). Table 5.1 below shows the estimated hazard of teacher attrition for each year that a teacher taught in North Carolina public schools.

5.1 Survival Analysis of Teacher Attrition

The initial analysis of teacher attrition showed that there were 10,093[30] initial teachers (or teachers at risk), and of those 257 quit in their first year, showing a hazard rate of 2.6 percent (shown below in Table 5.1). The hazard rate continued to increase through the second and third years of teaching and reached its highest rate in year 3 of 5.1 percent. Although the hazard rate dropped in the 4th year, it rose slightly in year 5 to a 4.5 percent chance of quitting teaching in public schools in North Carolina. The hazard then began a small yet steady decline in the remaining years. The cumulative hazard of teacher attrition for this study was 0.296, indicating a 30 percent hazard rate of a teacher leaving a public school by the teacher's 8th year. Additionally, I tested the linear relationship between the attrition rate and a teacher's year of experience, and there was no significant linear trend (as the coefficient was -0.00 SE=0.00).

[30] There were 150 first-year teachers in 2018 and were removed from this analysis given it was unknown if they left in their initial year, which reduced the number of teachers in the sample from 10,243 to 10,093.

Table 5.1. Hazard of Teacher Attrition by Years of Service

Teaching Interval	Initial Teachers	Number Quit	Hazard	Cumulative Hazard	Number Remaining
1	10,093	257	0.025	0.025	9,986
2	9,768	402	0.041	0.066	9,366
3	9,322	474	0.051	0.115	8,848
4	8,218	315	0.038	0.154	7,903
5	5,809	263	0.045	0.197	5,546
6	4,165	169	0.041	0.235	3,996
7	2,937	110	0.038	0.269	2,827
8	1,813	51	0.028	0.296	1,762
9	958	-	-	0.296-	958

One note about the trend of the hazard rate, is that it is abnormal compared to prior literature on teacher attrition. For instance, in a 2015 study of national teacher attrition rates in public schools, the US Department of Education, found that the hazard for teacher attrition reached a peak in the teacher's fifth year and then mostly stabilizes, with a substantially smaller hazard by year 6 (Gray & Taie, 2015). As the data for my study did not follow this pattern, I compared the significance of the log odds of each year by restricting the estimation to just those who had taught at least the specified number of years (i.e. four years or more). This is shown below in Table 5.2. From this analysis, the odds ratio was negative for every year of experience and was significant for every year. An additional observation showed that the odds of attrition increasingly decreased with each year of experience (except between years 3 and 4), ending with an odds ratio in year 7 of -0.78 (SE=0.14).

Table 5.2. Ratio of Elementary Teacher Attrition by Years of Service[31]

Year of Teaching	β	SE
2	-0.07***	0.02
3	-0.12***	0.02
4	-0.11***	0.02
5	-0.24***	0.04
6	-0.38***	0.07
7	-0.78***	0.14

To further examine if my data was an outlier compared to prior research, I also examined the odds ratios of teacher attrition by teacher characteristics to identify if teacher characteristics of teacher gender, race, or higher degree showed any differences in this data (shown below in Table 5.3). In these models, the hazard was highest for minority teachers and male teachers in their 3rd year of teaching with a hazard of 5.2 percent and 6.8 percent, respectively. However, for those with master's degrees, their hazard was highest in year 5 at 4.4 percent, though this may be the result of degree completion and more career opportunities. For all three of these subgroups the hazard after year 5 wasn't a steady decline, instead for minority teachers it increases in year 8 to 4.6 percent. There was also a larger drop-off for male teachers and those with master's degrees in year 8 (going from 4.5 percent to 2.5 percent for males, and 3.9 percent to 2.6 percent for those with master's degrees). Furthermore, the cumulative hazard was higher for minority and male teachers than that of all teachers in the study.

[31] Year 8 was not modeled given this was the highest year of service.

Table 5.3. Hazard of Teacher Attrition by Years of Service and Characteristics

Teaching Interval	Minority Teacher Hazard	Minority Teacher Cumulative Hazard	Male Teacher Hazard	Male Teacher Cumulative Hazard	Master's Degree Hazard	Master's Degree Cumulative Hazard
1	0.024	0.024	0.034	0.034	0.014	0.014
2	0.038	0.062	0.062	0.094	0.023	0.037
3	0.052	0.110	0.068	0.156	0.036	0.071
4	0.043	0.148	0.038	0.188	0.035	0.104
5	0.045	0.187	0.045	0.224	0.044	0.143
6	0.037	0.217	0.051	0.264	0.032	0.170
7	0.041	0.249	0.041	0.296	0.039	0.203
8	0.046	0.284	0.025	0.314	0.026	0.224
Beginning N	1,656		849		2,114	
Remaining N	137		61		341	

5.2 Analysis of Program Elements and Teacher Attrition

This next section of this analysis examined the relationship between covariates and the odds associated with a teacher's attrition using a logistic regression. I first examined the logistic odds of teacher attrition when covariates of school, teacher, and county unemployment were added (model 14, Table 5.4). After this initial model, I ran models with covariates of the institutions where the teacher was trained (model 15 and 16, Table 5.5), a logistic odds regression of program elements entered individually and in blocks by category (Table 5.6), and finally full logistic odds regression of teacher attrition and all identified elements within teacher training program (model 17 and 18, Table 5.7).

Before examining results from the logistic regression models, I ran a t-test between teacher attrition and independent variables to understand if any of the variables had a significant linear relationship to teacher attrition. However, no variables had a significant relationship to teacher attrition. I then examined the initial logistic regression of teacher attrition and covariates (shown below in Table 5.4). The variables tested in this analysis were time constant variables such as teacher gender, teacher race, and teachers with a master's degree. School variables such as high percentage of ESL students, high minority student population, high poverty, and county with high unemployment were not modeled as time varying in this study. Instead variables represent a 3-year average and were added as dichotomous dummy variables into the analysis (for instance high poverty =1 low poverty =0)[32].

Results showed that male teachers had a positive relationship to teacher attrition, with an odds ratio of leaving a public school in North Carolina estimated at 1.31 (SE=0.10, $p<0.001$), implying that the odds of leaving public schools increased 31 percent if the teacher was male, when controlling for all other covariates. Other teacher characteristics that were significantly related to teacher attrition were whether the teacher had a master's degree with an odds ratio of 0.72 (SE=0.04, $p<0.001$). Two significant school variables that were predictors for teacher attrition was the percentage of ESL and minority students at the school. For instance, the odd ration of teacher attrition at a school with a high percentage of ESL students was 1.21 (SE=0.06, $p<0.001$), whereas the odds ratio at a school with a high percentage of minority students was 1.74 (SE=0.09 $p<0.001$). This suggests that while the odds of a teacher leaving increased by 21 percent if there was a high percentage of ESL students in the school, they were even higher at 74 percent when the teacher taught at a school with a high minority student population, when controlling for all

[32] Covariates that are indicators of large percentages of identified subpopulations were defined by cut-off of the fourth quartile for each variable. A school with a high ESL population, is a school where 57 percent of the student population are ESL students; schools where a high population of students were economically disadvantaged, this is 64 percent of students at the school are identified as economically disadvantaged; schools with high percentages of minority students were those where 30 percent or more of students identified as minority students; and counties with high unemployment is any county with an unemployment rate of 4.7 percent or higher.

other variables. This finding is consistent with prior literature on school environments that are associated with high teacher attrition rates (Darling Hammond & Carver-Thomas, 2017). One additional variable that was significant at the $p<0.05$ level was teacher's race as an underrepresented minority, with odds ratios for leaving estimated at 0.84 (SE=0.06). Finally, results showed that employment opportunities within the county of the school where the teacher worked were mildly related to teacher attrition. For instance, the odds ratio for leaving public school teaching was 0.92 in counties with high unemployment, suggesting the odds of a teacher leaving in these counties was reduced by 8 percent when teaching in those areas. Although it was only significant at the $p<0.1$ level, this is consistent with literature that shows teacher attrition is tied to available employment opportunities (Goldhaber & Cowan, 2014).

Table 5.4. Logistic Odds Ratios for Teacher Attrition with Covariates

	Model 14	
	β	SE
Intercept	0.02	0.00
Year 2	1.71***	0.14
Year 3	2.14***	0.17
Year 4	1.60***	0.14
Year 5	1.97***	0.18
Year 6	1.77***	0.18
Year 7	1.65***	0.19
Year 8	1.24***	0.20
Male Teacher	1.31***	0.10
Master's Degree	0.72***	0.04
Minority Teacher	0.84**	0.06
High ESL School	1.21***	0.06
High Poverty School	0.96	0.05
High Minority School	1.74***	0.09
High Local Unemployment	0.92*	0.05
Pseudo R^2	0.02	
Number of Teachers	10,093	

***p< 0.001 level. **p< 0.05 level * p<0.1 level

5.2.1 Training University and Teacher Attrition

Once the relationship between teacher characteristics, school characteristics, and teacher attrition were determined, I examined whether the university where a teacher received his/her training had a relationship to the teacher's attrition. For this analysis there were 34 institutions[33] that were included (4 additional institutions than in the analysis of gain scores). My first step was to model university covariates without teacher and school covariates, and to then add in those covariates to isolate the relationship between universities and attrition. In model 14, the previous model without university covariates, the pseudo R^2 used to assess the proportion of variance explained by the covariates was small at 0.02. Additionally, in model 15 (the model with just universities) the pseudo R^2 was 0.01, while in model 16 (with school and teacher covariates as well as university covariates) the pseudo R^2 was 0.02. This small of an R^2 suggests the predictors added to these models do not define well the percentage of variance in teacher attrition.

As shown below in Table 5.5, six institutions had a significant odds ratio when entered into the model with only universities (model 15), however this was only at the $p<0.1$ level (and no institution was significant at a higher level in this model). The odds ratios between these institutions ranged from 0.50 to 1.87. Such wide differences indicated that some institutions reduced the odds of a teacher leaving the public school, while others increased the odds. For model 16, the odds ratio for university 9 became significant when all covariates were added at 1.55 (SE=0.47), though this was only at the $p<0.1$ level. Only 1 institution, university 6, was significant

[33] Institutions were included if there were more than 50 teachers in the dataset that were trained from the institution.

in both models with an odds ratio of 1.58 (SE=0.48) in the initial model, and then 1.64 (SE=0.50) in the full model. This suggests that training at this institution increased the odds of teacher attrition by 64 percent, when controlling for all other covariates.

Table 5.5. Logistic Odds Ratios for Teacher Attrition by Training University

	Base Line Model 15		Full Model 16		Teachers in Year 1	Teachers in Year 8
	β	SE	β	SE	N	N
University 1	1.12	0.32	1.10	0.32	1,589	316
University 2	1.13	0.33	1.13	0.33	1,339	185
University 3	1.54	0.45	1.45	0.42	1,033	159
University 4	1.14	0.33	1.08	0.32	825	179
University 5	1.32	0.38	1.34	0.39	827	131
University 6	1.58*	0.48	1.64*	0.50	357	44
University 7	0.98	0.36	1.00	0.36	123	17
University 8	0.96	0.29	0.97	0.30	490	72
University 9	1.50	0.45	1.55*	0.47	366	64
University 10	0.96	0.30	0.84	0.26	370	72
University 11	1.13	0.38	1.18	0.39	182	31
University 12	0.59*	0.21	0.65	0.23	211	44
University 13	1.31	0.43	1.26	0.42	184	26
University 14	1.31	0.42	1.07	0.35	195	40
University 15	1.61*	0.52	1.48	0.49	164	25
University 16	1.87*	0.69	1.68	0.62	76	3
University 17	0.81	0.30	0.76	0.28	128	22
University 18	1.14	0.38	0.85	0.29	187	29
University 19	1.54	0.51	1.46	0.49	132	28
University 20	0.88	0.32	0.65	0.25	136	17
University 21	1.08	0.42	1.09	0.43	76	18
University 22	1.28	0.47	1.25	0.46	94	14
University 23	0.92	0.33	0.86	0.32	132	24
University 24	0.81	0.32	0.87	0.34	94	20
University 25	0.84	0.34	0.88	0.36	82	15
University 26	1.15	0.44	0.96	0.38	86	11
University 27	1.29	0.45	1.38	0.48	119	16
University 28	1.32	0.47	1.21	0.43	97	25
University 29	0.50*	0.22	0.53	0.24	99	29
University 30	0.80	0.32	0.67	0.28	87	18
University 31	1.81*	0.70	1.39	0.54	53	15
University 32	0.95	0.39	1.00	0.41	68	20
University 33	1.00	0.42	1.05	0.44	63	14
University 34	1.33	0.53	1.24	0.50	60	11
Pseudo R^2	0.01		0.02			
Total					10,124	1,752

***p< 0.001 level. **p< 0.05 level * p<0.1 level

5.2.2 Logistic Regressions of Program Elements and Teacher Attrition

This section answers the third research question of this study by examining training elements tied to teacher attrition. For this section, I followed a similar analytical approach as in the student achievement models. I first entered the variables of program elements into the model individually, then in blocks, and then in a full model. Table 5.6 shows coefficients and standard errors for elements entered individually and in blocks, Table 5.7 shows the results for the full model with program elements. The institutions in this portion of analysis were the same as those included in the analysis of student gain scores, as those 24 institutions showed a significant effect on student gain scores and served as the basis for this additional analysis. For this sample, I retained over 60 percent of my original sample with 6,292 teachers in this analysis.

When training elements from curriculums were entered into the model individually, 20 elements were significant and 16 were significant at the $p<0.05$ level or higher. Additionally, when elements were entered into the models in blocks, there were often deviations from the individual findings. For instance, there were no elements that were significant from the individual analysis on faculty variables. However, in the blocked analysis, all variables except type of courses taught by adjuncts showed some significance. While number of FTE faculty and adjunct faculty were significant at the $p<0.001$ level, the size of the odds ratio was 1.01 (SE=0.00) and 1.00 (SE=0.00), respectively. This indicates these variables had a small effect on teacher attrition. The two remaining variables in this block were percentage of adjunct and faculty incentives for teaching, which were both slightly significant at the $p<0.1$ level with ratios of 0.89 (SE=0.07) and 0.81 (SE=0.10).

Elements within block 2 (Candidate Teaching Experiences) also showed similar changes, wherein only 1 variable was significant when entered individually (early teaching experience), though it was no longer significant in the blocked model. However, in the blocked model, there were 4 significant variables: classroom/teaching experience in intro course, number of courses with active teaching experiences, high number of courses with active teaching and hours of student teaching. For this section, having a classroom/teaching experience in the intro to education course had an odds ratio of 0.84 (SE=0.09), though at the p<0.1 level. The number of active teaching courses however, had a higher odds ratio of 1.13 (SE=0.07) at the p<0.05 level. When I examined this variable further with a dichotomous variable for high number of courses (15 courses) had a similar odds ratio of 1.13 (SE=0.09), though at the p<0.1 level. While the number of hours of student teaching was significant at the p<0.05 level, it had an odds ratio of 1.00 (SE=0.00), signaling no effect on teacher attrition.

Block 3, which contained a variable for university competitiveness and number of required courses, showed the most significant elements and variability within blocked models. For instance, variables for undergraduate SAT, as well as dichotomous variables for high (1090 or higher) and low SAT (below 900), were all significant in the individual and blocked analyses. When high and low SAT were entered into the blocked models, there were changes in the results and significance of 4 variables: number of math courses, number of teaching math courses, number of English courses, and number of foundation courses. To highlight these changes, I separated these results in Table 5.6, and results when high SAT was in the blocked model are reported under *Block 3A: Required Courses,* whereas results when low SAT was in the model are shown under *Block 3B: Required Courses.* When high SAT was in the model (block 3A), the number of English courses was significant at the p<0.05 level with an odds ratio of 1.22 (SE=0.11), as was the number of courses of teaching math with an odds ratio of 0.72 (SE=0.13), though at the p<0.1 level. Additionally, high SAT university has a strong relationship to teacher attrition with an odds ratio of 1.76 (SE=0.27) at $p<0.001$. However, when low SAT was in the model, the number of English courses and courses of teaching math were not significant, but the number of math courses were with an odds ratio of 1.18 (SE=0.09) and foundation courses had an odds ratio of 0.87 (SE=0.06) both at the $p<0.05$ level. Also, low SAT university has a negative relationship to teacher attrition with an odds ratio of 0.73 (SE=0.09), at $p<0.05$.

Elements in block 4 were consistent between both the individual and blocked models. For instance, the number of courses with imbedded instructional technology were significant at the $p<0.05$ level in both models with the highest odds ratio being 0.88 (SE=0.04) in the blocked model. When I examined this further, the variable for low number of courses (3 or fewer) was significant in both models as well, though changed in significance from $p<0.001$ (1.25 SE=0.08) in the individual model to $p<0.1$ (1.15 SE=0.08) in the blocked model. Additionally, the number of technology courses was significant at the $p<0.001$ level with an odds ratio of 0.88 (SE=0.03).

In block 5, all but one (course-based assessment) of the elements that examined methods to assess a candidate's teaching were significant. The first element to show a strong significant relationship was the number of courses that videotaped a candidate's teaching, with the same odds ratio in both iterations of 0.95 (SE=0.01) at the $p<0.001$ level. When I examined this further to test if a high (6 or more) or low (2 or fewer) number of courses had a relationship to teacher attrition, both were significant at the $p<0.05$ level. For instance, high number of courses had an odds ratio of 0.83 SE=0.05 and low numbers of courses had an odds ratio of 1.21 SE=0.08. While formal teaching assessment was only significant in the individual analysis (0.86 SE=0.06, $p<0.05$), the use of pre and post-tests was significant in both iterations. Indeed, this variable was significant at the $p<0.001$ level, and the odds ratio decreased from 0.73 (SE=0.06) in the individual analysis to 0.69 (SE=0.06) in the blocked model.

Finally, in the last two blocks fewer elements were shown to have a significant effect on teacher attrition. For instance, for elements pertaining to mentoring during the student teaching experience in block 6, no element was significant. For block 7, the elements that students, faculty, and program coordinators reported as notable program attributes diverged with findings from those in chapter 4. In the previous analysis, the variable noting candidate student teaching and candidate diversity were significant in the individual and blocked models, however, in this analysis, faculty as well as candidate preparation for state teaching exams were significant. The variable for faculty relationships had an odds ratio of 0.77 (SE=0.09) at the $p<0.05$ level and stayed significant in the blocked model, though at the $p<0.1$ level (with an odds ratio of 0.79 SE=0.10). Finally, candidate preparedness for state exams was similar in both iterations with an odds ratio of 1.13 (SE=0.08) when examined individually and 1.14 (SE=0.08) in the blocked model, both at the $p<0.1$ level.

Table 5.6. Individual and Blocked Logistic Odds of Attrition and Program Elements

	Individual Model		Blocked Model	
	β	SE	β	SE
Block 1: Faculty				
Number FTE Faculty	1.00	0.00	1.01***	0.00
Number Adjunct	1.00	0.00	1.00***	0.00
Percentage of Adjunct	1.02	0.03	0.89*	0.07
Adjunct Teach all Course Types	1.01	0.01	1.01	0.01
Faculty Incentives for Teaching	0.92	0.06	0.81*	0.10
Block 2: Candidate Teaching Experiences				
Early Student Teaching Experience	0.92*	0.04	1.11	0.13
Classroom Experience in Intro	0.93	0.07	0.84*	0.09
Number of Active Teaching Courses	1.08	0.06	1.13**	0.07
High No. Courses Active Teaching	1.10	0.08	1.13*	0.09
Low No. Courses Active Teaching	0.93	0.11	0.88	0.12
Computer Adaptive Teaching	1.02	0.06	1.03	0.10
Hours of Student Teaching	1.00	0.00	1.00**	0.00
Block 3A: Required Courses				
Average SAT Score (University)	1.00***	0.00	1.00***	0.00
High SAT University	1.33***	0.07	1.76***	0.27
Hours Prior to Student Teaching	1.00	0.00	1.00*	0.00
Number Method Courses	0.94**	0.03	1.03	0.08
Number Assessment Courses	1.08**	0.03	1.12	0.09

Table 5.6 Continued

	Individual Model		Blocked Model	
	β	SE	β	SE
Number Math Only Courses	1.17***	0.04	0.90	0.10
Number Teaching Math Courses	0.94*	0.04	0.72*	0.13
Number English Only Courses	0.94**	0.03	1.22**	0.11
Number Teaching Reading Courses	0.98	0.03	1.02	0.06
Number Diverse Learning Courses	1.02	0.03	1.02	0.07
Number Foundation Courses	0.89**	0.04	0.96	0.08
Block 3B: Required Courses				
Low SAT University	0.71***	0.07	0.73**	0.09
Hours Prior to Student Teaching	1.00	0.00	1.00	0.00
Number Method Courses	0.94**	0.03	1.00	0.07
Number Assessment Courses	1.08**	0.03	0.98	0.07
Number Math Only Courses	1.17***	0.04	1.18**	0.09
Number Teaching Math Courses	0.94*	0.04	0.81	0.14
Number English Only Courses	0.94**	0.03	1.02	0.08
Number Teaching Reading Courses	0.98	0.03	1.00	0.06
Number Diverse Learning Courses	1.02	0.03	1.11*	0.08
Number Foundation Courses	0.89**	0.04	0.87**	0.06
Block 4: Technology				
No. Courses with Technology Lessons	0.87**	0.04	0.88**	0.04
High Number of Courses with Tech	0.92	0.08	0.90	0.09
Low Number of Courses with Tech	1.25***	0.08	1.15*	0.08
Number Technology Courses	0.88***	0.03	0.82***	0.05
Block 5: Assessment of Teaching				
No. Courses that Record Teaching	0.95***	0.01	0.95***	0.01
High Number of Courses Record	0.88**	0.06	0.83**	0.05
Low Number of Courses Record	1.12*	0.08	1.21**	0.08
Formal Teaching Assessment	0.86**	0.06	1.02	0.08
Course Based Assessment	1.05	0.06	1.08	0.07
Pre-Post Student Tests in Review	0.73***	0.06	0.69***	0.06
Block 6: Mentor				
Mentor in In-Service Training	0.95	0.06	0.88	0.08
Mentor Review of Teaching Practice	1.02	0.06	1.11	0.09
Block 7: Noted Program Attributes				
Student Teaching Experience	1.10	0.11	1.06	0.12
Candidate Diversity	1.00	0.07	1.05	0.07
Faculty	0.77**	0.09	0.79*	0.10
Preparedness for State Exams	1.13*	0.08	1.14*	0.08

***p< 0.001 level. **p< 0.05 level * p<0.1 level

5.2.3 Full Model of Program Effects on Attrition

The last models tested in this analysis were model 17, a full model with all covariates and all program elements, and model 18 a full model with dichotomous variables for number of courses. Results for these models are shown below in Table 5.7. The variance in models 17 and 18 was 0.02 for both models. For this section, I selected the same variables that were significant in the previous analysis on student achievement in chapter 4. Using findings from the individual and blocked analysis, I also tested additional elements in the full model. These elements were: high/low SAT university, high number of courses (active teaching, instructional technology, and recording teaching), and the program attributes of candidate relationships with faculty and candidate preparedness for state exams. The variable for high SAT was not significant in any model, and therefore I only included the variable for low SAT university. When dichotomous variables for high number of courses were added, there were changes in the significance of program variables (those results are shown in model 18, Table 5.7). Additionally, when the two variables for program attributes were tested, the model showed signs of low power as the coefficients were large, yet were not significant, and were not included in the final model.

Two program elements were significant in model 17 at the $p<0.05$ level or higher: math only courses and number of courses that videotape candidate teaching practice. For the variable on the number of math only courses, this had an odds ratio of 1.34 (SE=0.17) at the $p<0.05$ level. In model 18, this element stayed significant at the $p<0.1$ level with a lower odds ratio of 1.22 (SE=0.16). Thus, for each additional course of math content, a teacher's odds of leaving public school teaching increases by 22- 34 percent. Next, the number of courses that record a candidate's teaching, was significant at the $p<0.001$ with an odds ratio of 0.92 (SE=0.02). The dichotomous variable for high number of courses that record a candidate's teaching in model 18, was significant at the $p<0.05$ level with an odds ratio of 0.71 SE=0.08.

There were two other variables that were significant in model 17: the number of method courses and the number of foundation courses. The number of method courses was slightly significant at the $p<0.1$ level with an odds ratio of 1.24 (0.17), but not significant in model 18. The number of foundation courses was also significant at the $p<0.1$ level in model 17 with an odds ratio of 0.85 (SE=0.08) and increased in model 18 to $p<0.05$ though with a smaller odds ratio of 0.76 (SE=0.07). Finally, although the variable for the number of courses that included instructional technology was not significant, when I tested the dichotomous variable for high numbers of courses in model 18, it was significant with an odds ratio of 1.41 SE=0.27 $p<0.1$. Thus, the higher the number of courses that included lessons on using instructional technology (10 or more), the teacher's risk of quitting increased by 41 percent.

Table 5.7. Logistic Odds Ratios of Teacher Attrition and All Program Elements

	Model 17		Model 18	
	β	**SE**	**β**	**SE**
Intercept	0.08**	0.08	0.04***	0.01
Year 2	1.75***	0.18	1.75***	0.18
Year 3	2.10***	0.21	2.10***	0.21
Year 4	1.58***	0.17	1.58***	0.17
Year 5	1.97***	0.22	1.97***	0.22
Year 6	1.52**	0.20	1.52**	0.20
Year 7	1.80***	0.26	1.79***	0.26
Year 8	1.18	0.24	1.18	0.23
Male Teacher	1.37***	0.13	1.37***	0.13
Master's Degree	0.66***	0.05	0.66***	0.05
Minority Teacher	0.91	0.08	0.91	0.08
High ESL School	1.13**	0.07	1.13**	0.07
High Poverty School	1.02	0.07	1.02	0.07
High Minority School	1.62***	0.11	1.62***	0.l1
High Local Unemployment	0.98	0.06	0.98	0.06
Faculty Incentives for Teaching	0.94	0.12	0.93	0.16
No. Active Teaching	1.08	0.20		
High No. Active Teaching			1.05	0.34

Table 5.7 Continued

	Model 17		Model 18	
	β	SE	β	SE
Low SAT University	0.96	0.15	0.80	0.13
Method Courses	1.24*	0.17	1.13	0.20
Math Only Courses	1.34**	0.15	1.22*	0.16
Teaching Math Courses	0.73*	0.15	0.69	0.18
English Only Courses	1.03	0.07	1.09	0.09
Teaching Reading Courses	1.03	0.07	1.04	0.07
Foundation Courses	0.85*	0.08	0.76**	0.07
No. Courses with Technology	0.96	0.09		
High No. Courses with Tech			1.41*	0.27
No. Courses that Record	0.92***	0.02		
High No. Courses Record			0.71**	0.08
Formal Teaching Assessment	1.28	0.30	1.02	0.23
Pre-Post Student Tests	0.98	0.20	0.97	0.20
R^2	0.02	0.02		
Number of Teachers	6,292			

***p< 0.001 level. **p< 0.05 level * p<0.1 level

6.0 Discussion and Implications

In this study I questioned, what, if anything, matters *within* teacher training programs as it relates to a student's achievement as well as a teacher's attrition. Prior literature has asked similar questions of teacher training by exploring the impact of programs themselves and these outcomes. Similarly, this examines the effects of teacher training programs on student and teacher outcomes. However, no previous studies have 'looked under the hood' and examined what elements exist within *elementary* training programs that may impact *elementary* student gains. Furthermore, while the literature on teacher attrition is extensive, only one study has ever examined elements of training programs and effects on teacher attrition (Ingersoll et al., 2014). To understand the research findings of this study, I walk through the results of each research question and examine what these findings mean in the context of the broader literature. I discuss the implications of these findings for the programs that provide the training, the districts that hire from those programs, and the teachers who will ultimately benefit from improved training. Finally, I consider future directions of this research, acknowledging this study is only one piece in understanding the complex nature of teacher training and its relationship to student achievement and teacher attrition.

6.1 Question 1

6.1.1 Elements of Program Quality in Elementary Teacher Training Programs

Before I could test the effects of different elements of teacher training programs on student and teacher outcomes, I had to first identify the elements that were part of elementary teacher training programs in North Carolina. To get at this answer, I examined program curricula, Title II reports from the US Department of Education, and I surveyed programs throughout the state. While I was interested in which elements were present in training programs broadly, I also wanted to know whether the teacher candidate experiences varied substantially depending on the program that trained them. The short answer is that yes, while on the surface training experiences are very similar, after digging deeper into 'how' a teacher candidate is trained, there was evidence of programmatic differences in the way candidates are trained.

To understand these findings it is important to note that in elementary education, the same teacher often teaches core subjects such as math and reading, and likewise, elementary training programs train teachers to teach all core subjects (Cohen, Ruzek, & Sandilos, 2018). For instance, in this study less than 30 percent of math teachers only taught math and not reading. This is especially relevant when comparing differences between program requirements, as differences in courses or other programmatic elements are likely due to survey samples[34] and not because of the subject the teacher taught. As such, findings in this discussion of program elements represent

[34] There were 16 total survey respondents, however 1 institution that responded was in the reading sample and not in the math sample.

overarching trends as well as unique elements present within all programs sampled, rather than differences by subject taught.

As noted previously, at first glance program requirements and training techniques are very similar. For instance, all programs have the same course components (such as pedagogy courses and subject specific courses). However, some programs have an emphasis on reading or math within the elementary teaching curriculum. All programs in the sample require approximately 2 courses in reading content (with a mean number of reading courses of 2.08), but the standard deviation is almost 1 course (0.88 in the reading sample), indicating some programs may only require 1 course on reading subject matter, whereas others may require 3. Additionally, this deviation is also present in the number of courses on teaching reading as the mean number of courses is 2.30 with a standard deviation of 0.97 in the math sample, though some programs may only require 1 course, yet others require more.

These trends are similar in math content and in courses on how to teach math as well. While programs require approximately 2 math content courses (with a mean of 1.87 courses), there are programs that only require only one math course (as the standard deviation in the math sample is 0.76). Additionally, for courses on how to teach math, programs require approximately 2 courses but once again, some programs only require 1 course on this topic (a mean of 1.74 courses of teaching math with a standard deviation of 0.81 in the math sample). This variation in subject content as well as content method courses is notable, as the literature on teacher training has suggested that increased training in subject content and subject teaching methods has an influence on a teacher's effectiveness (Henry, Campbell, et al., 2013). However, there is limited consistency amongst programs as each program has a different emphasis on content training (at least in the number of these required courses). One possible reason for such differences could be that the program identified a need in the community where most of their candidates teach (such as the need for stronger skills in elementary math). Another reason could be that while the program would like to have the same number of courses for each subject, programs are limited in the total program requirements and training hours they are able to require of a candidate and therefore cannot add more content courses with all the other requirements (Preston, 2017).

One additional finding in the review of program curricula was the limited number of required instructional technology courses. Among sampled programs, the mean number of courses was approximately 1 per program (with a mean of 0.83 courses of instructional technology and a standard deviation of 0.64 in the reading sample). This is of note as prior literature found that a higher number of instructional technology courses had a positive effect on a student's achievement, yet there are programs in this study that have no courses on instructional technology (Preston, 2017). That said, one reason for this could be that instead of having a single course, programs have embedded instructional technology in several courses throughout the training program. Indeed, findings from my survey show that of the required courses candidates take, on average 6 courses have instructional technology within them (though this had a wide range with a standard deviation of 2.67 in the reading sample). Thus, programs are indeed incorporating technology as part of their training, but this only became evident once I looked beyond broad program requirements.

Going beyond the formal structure of the curriculum, efforts to promote high quality instruction by faculty may be one way to improve teacher training. For instance, survey findings suggest that every program has at least one incentive to improve faculty teaching, though less than 50 percent of survey respondents have 3 or more incentives available to faculty (standard deviation was approximately 0.50 for both samples). This is relevant as there is evidence that faculty teaching techniques can garner improved student work products and improve targeted learning (Condon, Iverson, Manduca, Rutz, &Willett, 2016). Furthermore, Stieha, Shadle, &Paterson (2016) suggest that incentivizing faculty to adopt these practices can lead to improved faculty instruction overall. Although literature supports incentivizing faculty instruction, survey results show that programs vary in types of incentive. One reason may be the due to the very nature of teacher training programs themselves: faculty teach effective teaching methodologies and thus instructional expertise is expected.

Prior literature has suggested that student teaching experiences throughout the entirety of a candidate's training program, instead of just at the end of the training program, leads to the candidate feeling more prepared once they become the teacher of record (Ronfeldt, 2014). But just how much active teaching is present throughout teacher training programs in North Carolina? Compared to past models where candidates only taught a single class in their final semester/course, it has increased tenfold. For instance, approximately 75 percent of respondents noted there was an active teaching experience in the intro to teaching course. Furthermore, approximately 10 courses had active teaching experiences for candidates, though the standard deviation of 3.65 in the reading sample shows wide variation among respondents. While every program still has an official student teaching experience at the end of the training program, responses indicate that a candidate's teaching experience begins upon entry to the program and is no longer solely in the last/semester or year.

Finally, teacher preparation programs also vary considerably in how they assess a candidate's teaching. Prior literature has suggested that continued reflection and assessment of candidate teaching has been tied to improved teacher effectiveness (Chung, 2008; Ledwell & Oyler, 2016). One way that programs have added this continued reflection and assessment of the candidate is by videotaping the candidate's teaching and then reflecting on the candidate's practice, noting areas of improvement. Survey responses suggest that on average 3 courses require this of their candidates across programs, though the standard deviation is as high as 2.61 in the math sample. Another form of assessment of the candidate's teaching is a formal assessment, which provides a holistic review of the candidate's practice that is not tied to one course. When asked how many programs used this approach in candidate assessment, responses showed that 50 percent of programs provide assessment only in courses, and 25 percent of programs have an assessment process separate from courses. Although I did not ask follow-up questions about the assessments, programs may be using holistic assessments embedded within courses, which have been shown to be effective approaches to assessing and improving candidate teaching (Mason, Stanislawski, & Whitaker, 2019).

So, do candidate experiences vary between programs? Yes and no. Although all programs have the same general requirements, the implementation of those requirements on the ground varies widely. For instance, all programs have active teaching courses. While some programs have active teaching experiences in almost every course, others reserve these opportunities for only one or two courses outside of the student teaching experience. These results support the hypothesis presented in Preston's 2014 study on components of teacher training programs. In that study, she argued that differentiation shown in program curriculums was likely an incomplete portrait of training techniques among programs, and that the differences between programs were found in course assignments and candidate assessments. Survey findings in this study highlighted that argument as there was more variation around specific training practices than broad courses or overall training requirements.

6.2 Question 2

Before discussing the quantitative assessment of programs and their elements, it is important to note the limitations of the survey sample and its subsequent effects on the results of the analysis. Programs were selected to be surveyed if they showed a significant effect to student achievement in the initial quantitative analysis. This design was selected to provide a more qualitative understanding of programs through a review of their program elements. However, as shown in the results, this methodology reduced the total variance in the program elements that were tested in the analysis and subsequently the statistical power in the quantitative analysis of program elements was reduced. For instance, the total variance explained when program elements were added to the model was 0.02 in both reading and math. Given this limitation, the study findings are a more conservative assessment of program element effects than would otherwise be if the sampling approach included programs with a wider range of effectiveness.

6.2.1 Program Elements and Student Achievement

I considered a number of program elements, ranging from required courses to training techniques and teaching assessments within those courses. However, there was only one program element that had a significant relationship to student gains in both reading and math: low number of courses with instructional technology (3 or fewer courses). This element had a slight negative relationship in reading with a coefficient of -0.84 in multilevel model 10[35]. This contrasts with the

[35] Unless otherwise noted, in the discussion of program elements and student achievement gains, I primarily focus on full models, specifically multilevel models 9 and 10 and school fixed effects models 12 and 13 for both

math school fixed effects model, as low number of courses with instructional technology had a large positive relationship with a coefficient of 1.47 (model 13). Furthermore, when I tested a high number of courses of instructional technology (10 or more) in the reading model, there wasn't a significant relationship to student achievement. Thus, more than 3 courses, but less than 10 courses with instructional technology embedded throughout the curriculum are important for reading teachers in elementary grades.

These findings suggest that a higher number of courses that include instructional technology may be needed for those teaching reading than in math. However, as elementary teachers often teach both subjects, this finding likely highlights an issue within the training rather than number of courses: the application of the technology in different subjects. In elementary grades the use of instructional technology varies by subject. For instance, in elementary math teachers mostly use presentation technologies such as interactive whiteboards and document cameras to model and practice solving math problems (Polly, 2014). However, in reading, multiple technology strategies are employed, such as videos, interactive group writing exercises, shared documents for group reading, as well as tools to differentiate learning for each student (McDermott & Gormely, 2015). Although findings showed that the number of courses with instructional technology differs by subject, it is also possible that these results signal the importance of training on the practical applications of the instructional technology instead of the number of courses.

---

reading and math. These models include all covariates and identified program elements. All other models are baseline models, or were exploratory models used to explore relationships between student gains, universities, and/or program elements.

Given these differences in subject findings, I will now consider the program features specific to subject-matter that emerged in this study.

6.2.2 Reading

The results for the models in reading showed an interesting contrast: general teaching method courses had a negative effect on student achievement gains whereas courses of English-only content had a positive effect on student gains. In the multilevel model (model 9), general methods courses had a coefficient of -0.57 and a coefficient of -0.86 in model 10, thus suggesting that one standard deviation increase in method courses, would lead to a decrease in gains in a student's achievement in reading by -0.11 standard deviations. While it is counterintuitive that courses on 'how to teach' would have a negative relationship to student achievement, perhaps even more surprising is that the number of English-only content courses (such as English 101 or literature for reading teachers) had a significant positive effect in both multilevel models (in both models 9 and 10), and were further confirmed in model 12 with a coefficient of 0.50.

An initial read of these findings may cause some to question whether teachers even need to be trained in teaching methods, or do reading teachers just need training in English? Unfortunately, the answer is not a simple yes or no. To understand these findings and their complexities, I turn to the vast literature on teacher training. In 2005, in a study of teacher competencies after their training experience, authors Hill, Rowan, and Ball found that a teacher's knowledge of subject specific methods had a greater effect on student achievement than a teacher's knowledge of general teaching methods. Furthermore, researchers later found that a teacher's subject matter knowledge (such as knowledge in math or English content) was positively associated with a teacher's instruction and subsequently with student achievement (Ball, Thames, and Phelps, 2008; Hill, Ball, & Schilling, 2008). Researchers have thus argued that content knowledge is a prerequisite for teaching the content, though strong subject knowledge does not necessarily equate to strong subject teaching (Kleickmann et. al., 2013).

My results using the data from NC agreed with the existing literature in showing that general method courses had a negative relationship to student achievement, and content-specific courses in English matter. However, in contrast to the literature, courses on how to teach English did not have an effect on student learning gains in reading. While this deviates from the literature, it does not mean that general method courses should be removed, or that subject specific method courses do not matter, or even that candidates need more English courses. Indeed, it is likely that candidates receive an influx of general teaching methods in courses on content specific teaching courses and general method courses may be less useful and redundant. This suggests that programs may need to either reduce the number of general teaching method courses or modify courses on reading instruction so as not to over expose the candidate with training general teaching methods at the expense of training in reading methods or vice versa. Furthermore, as the literature shows stronger content knowledge is important for candidate development in teaching reading, programs should continue with these content specific courses as they modify other areas of the curriculum.

Moving beyond what is important for a candidate to learn within their training, I also examined whether it mattered how the candidate's teaching was assessed. Although no assessment method was significant in the full models, in the exploratory blocked models the use of student pre- and posttests was significant. That is, the use of elementary school student achievement data on tests to inform an assessment of the candidate's teaching practice had a positive effect on student achievement. Although there is scant research on this as an assessment tool in teacher training, some literature has suggested that using student learning (that is, learning of students taught by the candidate) as part of a candidate's student teaching experience can positively influence a candidate's teaching practice and ultimately the candidate's effectiveness when they are a teacher of record (Bastian, Henry, Pan, & Lys, 2016). Although this assessment method is currently present in only a few programs and was only effective in the blocked models, it does suggest one emergent area in the literature. Programs may want to consider adding this to their other methods of candidate teaching assessment to provide further information to candidates on the effects of their teaching.

6.2.3 Math

Turning now to findings specific to the models of program elements and math achievement gains, there were two elements that were significant: the number of foundations in education and a program's use of faculty incentives. The number of foundation courses showed almost a whole point increase in math gain scores at 0.77 in model 9 and in the school fixed effects model (the coefficient was 0.94). Since this was the only requirement of training elements that was significant, does this mean that teachers only need to be trained in educational foundations to have a positive effect on student learning in math? Definitely not. Although elements such as math content courses were not significant in this study, that does not mean that programs should stop requiring elementary teaching candidates to learn math content or how to teach math. Indeed, it is more likely that as Preston (2017) argued, covariation may exist between foundation and pedagogy courses in math and therefore these relationships are unable to be tested in the study (due to limited variation within program requirements). Furthermore, it could be that foundational courses like child psychology and child development are more necessary in math than in reading, given that a child's anxiety in math occurs early in a child's development (Ramirez, Chang, Maloney, Levine, & Beilock, 2015)

The second program element that showed a relationship to a student's achievement in math was the incentives for faculty teaching, with a coefficient of 0.64. Findings showed that one standard deviation increase in faculty incentives leads to an increase in math student achievement gains by 0.05 standard deviations. Although this element is not directly tied to the candidate's training, I included this element to test for the teaching quality of the faculty. Given that this is only significant in the math models, it could be a signal of the quality of the faculty teaching strategies, as elementary teachers are known to have higher math anxiety than teachers in other grades and therefore need differentiated training in math (even in math methods courses) compared to what is needed in reading (Van der Sandt & O'Brien, 2017). Additionally, this finding could be a broader signal that faculty are supported in multiple ways not just in their teaching. Such support may in turn lead to higher engagement of faculty, which may ultimately lead to candidate success (Stieha, Shadle, & Paterson, 2016).

6.2.4 Additional Observations

Lastly, I want to mention one element of teacher training that surprisingly showed no effect on a student's achievement in either subject: active teaching experiences. I tested several variables intended to capture the multitude of teaching experiences a candidate has throughout this/her training such as: the number of courses with active teaching opportunities, early teaching opportunities, as well as hours of student teaching. However, not one of these elements showed any significant relationship to student gains. I included these various elements to test whether more teaching opportunities for candidates has a positive relationship to teacher effectiveness. This begs the question once more: if there is no relationship to student achievement, should training programs remove this element from their training? Of course not. There could be many reasons these elements had no effects in my models. It could be that although every program has a different model of active teaching experiences, none of these models had an effect on candidate teaching. Another reason could also be that what really matters to a candidate's teaching development is not the total number of opportunities to teach a class, or when it occurs in the training sequence, or even the number of hours in a classroom. Instead, the quality of the experiences, combined with feedback on the candidate's teaching, could be more relevant to a candidate's development (Boyd et al., 2009; König, 2013; Ronfeldt & Reininger, 2012).

Although the focus of this study is to test the effects of training on all students, I performed an exploratory analysis of the effects of training programs on student achievement within schools that had high percentages of low-income students[36]. The results of this analysis for both reading and math is shown in Appendix C, Tables C.1(reading) and C.2 (math). One initial observation was that programs that were statistically significant and effective among all teachers had similar coefficients but were not significant (which may be due to smaller sample sizes). For instance, 20 universities were significant in the reading multilevel model for all schools (reading model 6) yet, for high poverty schools, the multilevel model (model 6) showed only 6 universities that were significant. Given this was only exploratory, I did not pursue a rigorous analysis to compare coefficients between models. However, as an informal comparison of program effectiveness among all teachers compared to those in high-poverty schools, I considered the number of

[36] Schools with high percentages of low-income students were identified as such if 64 percent or more students at the school were economically disadvantaged (determined from the top quartile of school poverty).

programs that appeared in the bottom and top quartiles of program effectiveness rankings for all schools, but not for high-poverty schools (sorted by coefficient).

Generally, programs that were in the top for all schools, were in the top for high-poverty schools. However, as has been shown throughout this study, there are variations between reading and math. For instance, there were minor deviations from programs in the top quartile for programs in the multilevel reading model as University 25 moved to the first rank, instead of fifth (as it was in the all school analysis), and University 15 and 9 also moved up from 10^{th} to 5^{th} and 6^{th} (respectively). In the bottom quartile for reading there was more movement. The biggest change was that University 24 was in the top 10 for all programs, but dropped to 25^{th} when only high poverty schools were examined. Similarly, University 28 was in the middle at 14^{th} for the all school analysis, however, dropped to the last program for high poverty schools. Thus, for reading, trends were similar to the initial analysis, though there were notable shifts in top and bottom programs rankings.

For the math multilevel models (model 6), there was more movement between the two analyses than was seen in the reading model. In the top quartile for all schools University 21 was first, yet dropped to 8^{th} for the high poverty analysis. Furthermore, only University 19, 10, 13 and 23 remained in the top quartile in the high poverty analysis, and the rank of those programs also changed. For the bottom quartile, the rank of programs in the multilevel math model for high poverty schools also were different from that of all schools. For instance, all ranks changed, as University 29 was 24^{th}, but in the high poverty analysis it dropped to the last program (with a large negative and significant coefficient). However, three universities also moved from the middle in the all school analysis to the bottom quartile in the high poverty analysis (university 2,16, and 18). As noted previously, the exploratory models into program effects on student achievement in high poverty schools had few significant findings. However, the movement of programs in this analysis suggests that there is an interaction between training and high poverty schools and warrants further investigation into the relationship between how a teacher is trained and subsequent effects on low-income student achievement.

6.3 Question 3

6.3.1 Teacher Attrition in North Carolina

Although one goal of teacher training programs is to train candidates to teach students, they also prepare candidates for the profession itself. Thus, to assess a more holistic measure of a program's effects, I examined whether there were real differences in teacher attrition rates between programs. My data showed that when a new cohort of teachers graduate from college and begin their first assignment, there is approximately a 3 percent drop-off in teachers by the end of their first year. This drop-off peaks in the teacher's third year of teaching to 5 percent, though it does not begin a real decline until after the fifth year, followed by incremental declines in subsequent years. This differs from national trends in teacher attrition though, as attrition rates peak in the teacher's fifth year and then substantially decline by year 6 (Gray & Taie, 2015). I ran several analyses to test if my data was an anomaly compared to national trends and concluded that although my results did not follow national trends, it did follow attrition trends for teachers in North Carolina. Specifically, attrition data for North Carolina show that rates are much higher in the first 5 years of a teacher's career (and even higher in the first 3 years) but there is a gradual drop after year 5, much like in my data (NCDPI, 2018b).

What is unique about teaching in North Carolina that may lead to this deviation in attrition rates? While several factors influence teacher attrition, state context shows that state finances and policy changes have likely played a role in attrition trends. For instance, in a study of teacher attrition in North Carolina, authors Clotfelter, Glennie, Ladd, & Vigdor (2008), found that teacher attrition in North Carolina could be reduced by 17 percent, even in experienced teachers, through financial increases and bonuses. Although this is not necessarily unique to teachers in North Carolina, teacher salaries in North Carolina have actually decreased by almost 12 percent since 2000 (accounting for inflation), which is substantially larger than the national decrease of 2 percent (Rezsonya, Yu, & Drew, 2019). Thus, one potential reason for this steady decline is that teacher salaries are not keeping up with inflation and as teacher experience increases, they may be able to find more lucrative employment opportunities elsewhere. Furthermore, attrition is highest for teachers in charter schools, as they have fewer teaching options (given they do not hold licenses) and their salaries are lower. Although charter teachers were not included in this study, teaches within this sector may be contributing to overall attrition rates of public school teachers within the state of North Carolina. Unlike most states, another concern that may influence teacher attrition in North Carolina is that in 2014 the state changed the law so that teachers no longer received a financial increase in salary if they received a master's degree. Prior to this law, teachers gained a 10 percent increase in salary when they received a master's degree (NCDPI, 2020c). This could have an effect on teacher attrition as the teacher may still pursue a master's degree with the intent to leave the classroom and move into educational leadership or other fields (as master's degrees take 2-3 years to complete such movement may begin in year 6 or later in the teacher's career).

One last note about teacher attrition in North Carolina is that in a 2016-2017 survey of teachers who quit teaching in the state, the top third reason teachers gave for quitting was that they wanted to leave the profession completely (10.6 percent of respondents). The only reasons above quitting the profession entirely were family relocation and retirement. Approximately 45 percent of those respondents who wanted to leave teaching were within the first five years of their teaching career (NCDPI, 2018b). This signals two things: more than half of teachers who decided they wanted another career did so after 5 years of teaching, and a substantial portion left because they did not like teaching. This is different from dissatisfaction with teaching, which the survey also asked respondents, and those unhappy with teaching accounted for less than 2 percent of respondents. Thus, North Carolina is unique in that either due to lower salaries or desire for other careers (or a combination), teachers more regularly leave after their fifth year.

6.3.2 Teacher Training Programs and Teacher Attrition

Before examining effects of program elements on a teacher's attrition, I first asked whether attending a specific university (i.e. training at program A versus program B) had an effect on a teacher's career in public schools. Results from model 16 showed that training at two institutions (University 6 and University 9) increased the chances a teacher would leave public schools. Interestingly though, in the analysis on student achievement both University 6 and University 9 had a positive association with student achievement in both reading and math. This suggests that teachers trained at these institutions had positive effects on student achievement gains, though were also more likely to leave public teaching in the state. This diverges from overall findings on teacher attrition which suggests that more effective teachers are more likely to stay teachers longer (Goldhaber, Gross, & Player, 2011; Papay, Bacher-Hicks, Page, & Marinell, 2017). These findings also deviate from the literature on training program effects and teacher attrition. For instance, Goldhaber and Cowan (2014) found there was no pattern between programs with high attrition rates and teacher effectiveness as those with the highest attrition rates had both the highest and lowest effects on student achievement. They argued that if more candidates were to be trained at the most effective programs this likely would not have an impact on teacher attrition. However, findings from my study suggest that there is a trade-off between teacher effectiveness and attrition, with the most effective teachers leaving the career earlier than other teachers. That said, as this was not a primary focus of the study, I did not explore this relationship at the teacher level. As such, any trade-offs seen between these two outcomes in North Carolina are not fully understood at this time.

6.3.3 Program Elements that Increase Teacher Attrition

Similar to the analysis on student achievement, I considered effects of program elements in a teacher's attrition that ranged from required courses to training techniques used within those courses. This section will first discuss the program elements that were found to increase a teacher's attrition and then examine the elements that reduced the chances of a teacher's attrition. The elements that were found to have a positive relationship to attrition (i.e. elements that increase a teacher's chances of quitting the profession) are: number of mathematics courses, number of general teaching methods courses, and number of instructional technology courses. I found that for each additional course of math content (thus math-only courses and not courses on how to teach math), a teacher's odds of leaving public school teaching increases by 34 percent. As discussed earlier in question 1, some programs have a different focus on requirements and, although elementary programs train candidates in all subjects, some programs may only require 1 math course whereas others may require up to 3 (the average across programs is 1.88 courses in math content, with a standard deviation of 0.74). While elementary education programs were the only ones included in this study (and thus subject-specific degrees such as math were excluded), this is similar to findings on attrition and training of math and science teachers. Ingersoll, et al., (2014) found that elementary teachers with degrees in math and science had higher attrition rates, and argued these teachers tended to leave the profession because although their command of the subject was strong, their pedagogical knowledge of teaching math was weaker. It is possible that this is happening in my data as well, as candidates that attend programs that have a focus on math may focus more on the math content rather than the methods courses and are then less comfortable with the pedagogical aspects of teaching math.

Another element that led to an increase in a teacher's attrition was the number of general method courses. Findings showed that for every additional course in general teaching methods, a teacher's odds of leaving the profession increases by 24 percent. Although limited research has discussed the effects of general method courses on a teacher's attrition, author Ronfeldt (2014) found that attrition rates decreased when higher numbers of method courses *and* hours in student teaching were combined. Their findings suggest that method classes alone are not responsible for teacher attrition, and it is likely the interaction between method courses and student teaching led to a reduction in attrition, instead of general method courses themselves. As my study examined the effects of method courses when controlling for all other program elements, it may not necessarily signal that method courses increase teacher attrition, but that for method courses to be effective in retaining teachers, they should be paired with student teaching experiences.

Lastly, this study showed that the higher the number of courses that included lessons on using instructional technology (10 or more), the higher the teacher's risk of leaving teaching early (a 41 percent increase). While the literature is scant as to why this might be, it is possible that as a teacher's skill in using instructional technology increases, the teacher's employment opportunities outside of teaching also increases. Additionally, more exposure to instructional technology may also create more career paths in education that were not examined in this study, such as in areas of curriculum or instructional design, or instructional technology experts.

6.3.4 Program Elements that Decrease Teacher Attrition

While program elements noted above increased a candidate's likelihood of quitting teaching, three elements were shown to reduce a candidate's chance of leaving: number of courses that record a teacher's teaching, number of foundation in education courses, and use of pre-and posttests in a candidate's teaching assessment. Findings showed that a teacher's odds of leaving public school teaching decreases by 8 percent for every increase in the number of courses that record a candidate's practice. Additionally, as the number of foundation courses increased, there was a 24 percent decrease in a teacher's attrition. These findings align well with those of Ingersoll et al., (2014) who found that teaching methods alone were not enough in candidate training, instead candidates needed feedback on their practice coupled with foundational education courses to reduce a teacher's attrition. Authors argued that continued feedback decreases a teacher's attrition by 58 percent and courses in education foundations decrease attrition by 51 percent. This was consistent in my study: although recording the candidate's practice is not necessarily a formal review practice of the candidate's teaching, it is evidence of continued review of one's practice. Thus, findings further suggest that feedback on a candidate's practice coupled with courses on student learning can reduce teacher turnover.

The final program element that showed a decrease in teacher attrition was the use of student pre- and posttests in candidate reviews. Findings showed a 31 percent reduction in a teacher's chances of leaving the public-school system when their training included these measures in teaching reviews. While the research is limited on effects of this method of candidate assessment and teacher attrition, as discussed earlier in findings of question 2, research suggests that using student learning in candidate assessments can lead to a more effective teacher. Additionally, this is the single program element that is positively tied to teacher effectiveness and decreased attrition. However, as it was only significant in blocked models (for both analyses), this may only be relevant in comparison to other assessment methods.

6.3.5 Institutional Selectivity and Teacher Attrition

Although the main findings from the study were from the full model, it is worth mentioning an element that was significant in the blocked model: the competitiveness (highly competitive or not competitive) of the university where the teacher was trained. This was significant in the blocked models when entered with required courses. The dichotomous variable for high SAT score[37] university indicated a strong positive relationship with teacher attrition (odds ratio of 1.76), whereas the dichotomous variable for low SAT score university had a negative relationship to teacher attrition with an odds ratio of 0.73. This is consistent with Kelly and Northrop's (2015) study on graduates from highly selective institutions which found that those graduates were 85 percent more likely to leave the teaching profession, compared to those from less selective institutions. This finding, as well as prior literature, suggests that teacher attrition may be more influenced by overarching structural and compositional features of the university where the teacher was trained rather than elements within.

---

[37] High SAT is 1090 or higher and low average SAT is below 900.

6.4 Regional Context

Before exploring what these results may mean for teacher training programs and the broader community, it is important to discuss the North Carolina context as it relates to my findings as well as teacher training in general. North Carolina has long been known as a leader in education reform across all of k-12 education, however most recently this has been due to policies in teacher training (Henry et al., 2014). These reform efforts are the result of many forces specific to the state itself, but one of these reasons is likely because the variety of training programs in the state allows for such experimentation. For instance, the state has approximately 46 total traditional training providers (the number of universities that offer teacher training), yet comparatively the median number of providers in the US is 19 per state. Regionally, North Carolina is a leader in the number of providers as the neighboring state of South Carolina has 31 providers, Georgia has 40, Virginia and Tennessee have 36 providers, and the more populous state of Florida also has 46. North Carolina is not just a leader in traditional providers of teacher training, but also has the most alternative training programs in the US. Furthermore, due to the multitude of opportunities to train teachers in North Carolina, the state produced one of the largest numbers of graduates from teacher training programs in the US in 2016 (US Department of Education, 2016c).

Although this study is primarily about teacher training programs, it is also necessary to give some context of the teaching profession itself in North Carolina. As of 2019, the state had the ninth largest population of elementary education teachers in the US (for reference, it is also the ninth most populous state in the country) (US Department of Labor, 2019). That said, annual mean wages ($48,560) for elementary teachers in the state are 47th in the US when compared to wages for elementary teachers in all other states (and DC). However, the mean wage amount masks the wide geographic differences in teacher wages throughout the state. While most salaries are between $43,000-$50,000. the mean salary in one large coastal city is over $70,000 and the mean salary in a rural city within the Appalachian region is approximately $41,000. Such disparities are often due to local salary supplements on top of the base salary wages set by the state, which are funded through local income taxes (NCDPI, 2020c). That said, the state has implemented a merit-based pay system, which provides a bonus to 4th and 5th grade reading teachers and 4th-8th grade math teachers who raise student achievement scores (NCDPI, 2020c). Although there are state bonuses that could increase a teacher's wages, amounts are not enough to make up the difference between the lowest and highest earning cities. Furthermore, while there are additional add-ons in the state such as principal pay, teaching AP classes, and National Board Certification, the state recently eliminated pay increases for those with master's degrees, thus reducing the financial benefits of further education. Although the state has one of the largest populations of elementary teachers in the US, the salary differentials both compared to national wages as well as throughout the state, signals that the local context likely dictates whether it is a desired career path for people to pursue and for teachers to stay long-term.

Such disparities in teacher salaries are not the only disparities present in education in North Carolina, however. The state also has a diverse population, in which there are striking inequalities between schools and districts. For context, as of 2016, 51 percent of students enrolled in k-12 schools in North Carolina were white (non-Latino) students, 29 percent were black students, and 16 percent were Latino. However, these percentages have been changing over the last 10 years as the population of white students has slightly decreased (from 58 percent in 2006), the Latino population in k-12 schools has doubled since 2006 (such trends are similar in both traditional public and charter schools) (Clotfelter, Hemelt, Ladd, & Turaeva, 2019).

When examining enrollment patterns of students by race and economic status, Clotfelter, et al. (2019) found that a large proportion of schools throughout North Carolina were both racially and economically segregated.[38] For instance, approximately 20 percent of schools were racially segregated, and similarly19 percent of schools were economically segregated. Researchers also found that although the state population itself is racially and economically segregated geographically, school segregation was evident in all geographic regions. While racial minority populations are geographically segregated from white populations (and low-income populations from high-income populations), student segregation occurs among individual schools within all school districts. Thus, segregation is not limited to one region of the state, but instead is dependent on the school itself (Clotfelter, et al., 2019). Clotfelter et al., (2019a) further examined these disparities in racial segregation and determined these patterns were present in both traditional public schools as well as charter schools. Individual charter schools initially were not racially segregated (even in geographic regions that were less racially diverse), however as charters have grown throughout the state, they have followed these trends as individual schools are highly segregated throughout the state.

What do these segregation patterns mean for education overall in the state? One issue is that these patterns are heavily tied to racial achievement gaps as well as lower teacher quality. For instance, as Reardon (2016) argues, racial achievement gaps are more correlated with school segregation than with geographic segregation. Furthermore, the author's results show that high-poverty schools are less effective at raising student achievement than those with lower percentages

[38] For this section, segregation is measured as the proportional gap between the actual and maximum extent possible of contact between two groups of students (such as white and non-white or poor and non-poor).

of low-income students. This is, in many ways, tied to lower teacher quality at schools with more racial minorities as well as more impoverished students. As Clotfelter, Ladd, Vigdor, and Wheeler (2007) noted, in high poverty schools throughout North Carolina principals and teachers alike have lower qualifications, and subsequently student achievement in both math and reading are lower. While this similar for schools with high percentages of racial minorities, Fahle, Reardon, Kalogrides, Weathers, & Jang (2020) found these schools also have higher teacher attrition (and thus more entry level teachers) and higher percentages of teacher absences, all of which are tied to lower student achievement.

The second reason it is important to note the impact of segregated students throughout schools in North Carolina is that the state has adopted several policies and programs to try and direct high-quality teachers to schools with the greatest need. One such policy is the North Carolina Teaching Fellow Program. In 2017, the state reestablished this program with the goal of increasing the pipeline of highly qualified teachers to low-performing schools and high need subjects (especially STEM and special education) (North Carolina General Assembly, 2017b). This program recruits prospective teachers based on their academic merit and then, once the prospective teacher graduates, the state will pay off a candidate's loans (up to $8,250 a year) for their training. To be eligible, the prospective teacher must attend one of five participating programs in North Carolina (North Carolina General Assembly, 2017b). Although this is only one example, it highlights how state initiatives to address inequalities can also impact teacher training throughout the state.

As noted earlier, North Carolina is not only a state with large inequalities in the education system, but it has also been a leader in education reform throughout k-12 and in teacher education. In addition to the North Carolina Teaching Fellow Program, the state has modified training program requirements (such as the implementation of edTPA), they have increased alternative routes to teaching and adjusted policies to support these routes, and they have also published ratings of training programs, to name a few (Henry et al., 2014). Given this state context, I questioned whether there were any noticeable effects in the overall teacher workforce from these efforts. Although this study was not necessarily about policy and reform efforts themselves, there are signals that there are both positive as well as negative changes to the teacher labor force. The first signal was that the teaching workforce is overwhelmingly white and, although there are programs to increase high quality teachers across the states, these findings do not suggest there is a focus on increasing racial diversity of teachers. This is evident in both traditional public and charter schools. Furthermore, similar to k-12 schools, training programs are heavily segregated, with white teachers predominantly trained in certain programs and candidates that identify as racial minorities in other programs. Another area of concern is that attrition rates are not stable after the 5th year of teaching and schools with high percentages of minority students are the number one predictor of teacher attrition (in every model). Thus, while reform efforts have tried to address teacher quality, they have not succeeded at keeping teachers in public classrooms and have struggled even more to keep teachers in racially diverse classrooms. However, there are almost no differences in rates by program, suggesting attrition is less tied to the program itself and that state policies or other school elements could be larger contributors to teacher attrition (as discussed in the analysis on teacher attrition).

Lastly, results from this study show that some programs are more effective at raising student achievement scores than others. Although program effects are approximately less than a one-point increase in a student's score, it suggests that the great attention to program reform could increase teacher quality. However, while this attention appears to be moving teacher quality in the right direction, the overwhelming concerns about state reforms to programs that were heard in the interviews in this study cannot be ignored. As such, while programs may show positive effects to student achievement, the pressure programs noted due to these reforms could have lasting negative effects on a candidate's development that are not measurable in this study.

6.5 Implications

6.5.1 Implications for North Carolina Programs in Study Sample

After discussing the results of each research question, an even bigger question remains: what do these findings actually mean for the training programs that were examined in this study? Survey responses as well as the vast literature on teacher training have repeatedly emphasized that programs face substantial pressure in training teaching candidates to be more effective as soon as they enter their first classroom (Goldhaber, 2019). However, as these programs work to respond to pressures, they are often limited by programmatic and financial constraints. These limitations, when coupled with the lack of sufficient evidence as to what 'matters' in teacher training, further complicates the ability of training programs to respond to the demands of the teaching workforce (Goldhaber & Ronfeldt, 2020). This is not to say that there is nothing programs can or should do. Indeed, there is evidence that what some programs are doing is effective and could be further tested among all or a subset of additional programs. Based on the results from this study, I have outlined three areas that the elementary teacher training programs that were examined in this study may want to focus their attention when reviewing training elements within their program. Although the programs included in the study sample do teach a majority of teaching candidates in the state, data limitations do not permit valid inferences about programs not included in the study. As such, references to programs and the implications study findings may have on programs refer solely to the programs that were examined in this study.

6.5.2 Implications for Sampled Programs: Course Requirements

Should teacher training programs included in this study de-emphasize courses that prepare teachers in general pedagogical methods and place greater emphasis on subject-specific pedagogy? Although the initial takeaway might be that teachers only need training in specific subjects, findings suggest a less direct approach. I found that more courses on teaching methods decreased student achievement in reading and increased a teacher's chances of leaving the profession. However, this does not suggest removing general teaching courses. Course requirements that were tested in this study were represented by the number of courses within categories (such as methods), and as such findings are tied to whether more or less of a course within that category is needed. Thus, although findings suggest reducing the number of method courses in reading, it does not mean eliminating them altogether. Instead, programs could consider modifying the number of general teaching method courses required. Furthermore, while subject-specific training courses did not show any effect on student achievement or a teacher's attrition, it is likely that these course categories (general teaching and teaching a specific subject) complement one another in that method courses reiterate lessons learned in subject-specific method courses, and vice-versa. Therefore, if a program were to alter the number of general method courses, they would likely want to do that simultaneously with alterations to the course curriculum, or requirements to subject-specific method courses.

6.5.3 Implications for Sampled Programs: Training Within Courses

Another observation is that what likely matters in training effective teachers who stay in the profession is not the number of credit hours or number of method courses. Instead, what happens *within* courses is likely more important. I found that among the programs in the study the number of courses that included instructional technology had an effect on student achievement and teacher attrition. This was also the case for the number of courses that record a candidate's teaching, which led to a decrease in a teacher's attrition. These findings are an indication that what matters inside the training is at least as important as the course topic itself for the programs sampled. However, this does present a challenge for those programs, as adding or removing specific courses (or categories of courses) is easier than altering training techniques within a course. For instance, standardizing course categories to include across programs is easier than identifying courses in which to include lessons on instructional technology, or which courses should record candidate teaching (and the training associated with those recordings). While modifications to elements within individual programs may be more difficult to add across programs, sampled programs may want to further investigate the effectiveness of the training techniques and methods that occur inside their courses rather than the courses themselves.

6.5.4 Implications for Sampled Programs: Student Achievement and Teacher Attrition

Perhaps the most frustrating finding for programs in this study is that elements that increased a student's achievement were not necessarily the same as those that decreased a teacher's attrition. Does that imply that programs must choose between student achievement or attrition when deciding on elements to include in their training? Of course not. Instead, when these programs revise offerings and techniques, they should focus on the 2 elements that showed a strong effect in both analyses: number of foundation courses and the number of method courses. Furthermore, while a low number of courses that included instructional technology were associated with gains in reading achievement, a high number of courses with instructional technology indicated a higher likelihood of teacher attrition. Although low numbers of courses with technology were unrelated to patterns of attrition, these findings together (for achievement and attrition) may suggest that more than 3 but less than 10 courses that include instructional technology could lead to lower teacher attrition. Given these findings, the programs studied may want to explore these courses as well as program offerings within courses as it could improve teacher effectiveness and teacher attrition.

6.5.5 Implications for Hiring Districts and Public Schools in North Carolina

Districts, and the schools where candidates ultimately work, have a strong interest in the quality of the programs where their teachers are trained. Although this is primarily because districts want to hire teachers who are effective in the classroom, prior literature suggests that teachers tend to apply to (and subsequently work in) areas where either the teacher themselves went to school or where they were trained (Killeen, Loeb, & Williams, 2015; Ronfeldt & Reininger, 2012). Authors Goldhaber, Krieg and Theobald (2016) also found that where the candidate completed their student teaching experience (which is often influenced by the location of the preparation program) was similarly predictive of where the candidate will ultimately have their first teaching position. Furthermore, while several factors influence why a principal ultimately hires a teacher, principals and hiring districts tend to hire from a program with which they are familiar (Boyd, Lankford, Loeb, Ronfeldt, & Wyckoff, 2011; Rutledge, et al., 2008). This literature suggests that employment and hiring decisions are often locational and dictated by factors outside the quality of the training itself.

To date, there is limited evidence that principals consider training program quality in their hiring decisions (Goff, et al., 2020). However, there is evidence that some training programs have a weak (but significant) relationship to teacher effectiveness and attrition (Goldhaber, 2019). Furthermore, findings in my study suggest that beyond the sampled programs themselves, there are some programmatic elements that have a relationship to achievement and attrition. Given the strong negative effects of high teacher attrition, which has been shown to be both costly and a negative effect on student achievement, districts may want to expand recruitment efforts to specific programs as well as those that include elements that have positive relationships to teacher effectiveness and teacher retention (Goldhaber & Cowan, 2014). Although Goldhaber & Cowan (2014) did not test this, they argued that one benefit of increasing recruitment of teachers from specific programs (those tied to lower attrition and higher student achievement), would be that districts could improve their overall candidate pool which could translate increased student and financial gains.

6.5.6 Implications for Teaching Candidates in North Carolina

Finally, this study provides implications for teacher candidates themselves, who will be trained by sampled programs in NC and may ultimately become teachers. While this study does not necessarily investigate pre-service teachers' individual experiences within the program, presumably teachers encounter real effects of program quality in their ability to be successful with students. Indeed, states (including North Carolina) have begun to require programs to provide more data on graduate outcomes such as attrition rates and achievement scores so that prospective teachers can make more informed decisions between programs. However, is this enough information? Are there additional indicators prospective teachers should consider when choosing their training program? Absolutely. While candidates want to know these basic outcome measures, they may also want to know how satisfied principals are with graduates of respective programs. One way of measuring this is through teacher evaluations that include information beyond student achievement scores, such as ratings on content knowledge, classroom management, and lesson design. Thus, providing information on teacher evaluations linked to programs can signal to prospective teachers areas of potential weakness that graduate programs may experience (Bastian, Patterson, and Pan, 2018).

Future candidates may also want to know about graduate teaching efficacy. That is, the graduate's confidence in both motivating students and raising student achievement once they become teachers. Although this is a more qualitative metric, teacher confidence has been linked to small increases in student achievement and can likewise signal to a candidate one of the strengths of the program (Kim & Seo, 2018). As prospective candidates decide between training programs, data on student achievement and attrition are only two pieces of what makes an effective teacher. To more fully inform candidates' decisions, programs and states may want to provide candidates with anonymized teacher evaluations of graduates, as well as the ratings of teacher efficacy of program graduates, so that candidates can more readily understand the strengths and weaknesses of the training they will receive.

6.6 Design Considerations for Further Research

Research on teaching training rarely examines elements that are unique to individual programs. Instead, it focuses on one individual element across programs or the program as a whole. One reason for this could be the difficulty of distinguishing between programs and identifying elements of effectiveness *within* programs. As teacher training programs across the US and notably in North Carolina move toward more standardization, identifying elements unique to programs is becoming even more difficult. However, the significant findings from this study suggest there are meaningful differences between programs in their training elements. Below, I discuss different approaches to the study of teacher education program effectiveness that can identify program elements as well as determine whether those elements have a relationship to achievement and/or teacher attrition.

Although I applied multiple methods to identify elements distinct to individual programs, I was not able to capture all elements. I examined program curriculums and surveyed programs thinking that any differences would be clearly identifiable. However, program offerings and techniques that occur within the program are much more nuanced than can be captured in a survey or review of program documents. Furthermore, surveys relied on voluntary participation and responses were only based on the respondent's interpretations and perspective. Thus, one limitation that is in all survey research, is that responses only showed what respondents themselves thought were unique to their programs. Additionally, I only surveyed program coordinators and did not include candidates or the faculty who are doing the training. While I found more elements unique to programs than in other studies, this study did not fully capture all elements that may matter in a candidate's training. One research method that could be applied to address this is a case study of several universities that would include course and training observations as well as interviews with faculty, candidates, and program coordinators. Findings from these case studies could then be tied to student achievement and attrition rates from program graduates. While this study design may not allow for the inclusion of all programs in the state, it could provide robust documentation of elements within programs that cannot be easily identified through surveys or public resources and may influence candidate outcomes.

One of the limitations in this study is related to the characteristics of the data available to answer my research questions. For instance, I was not able to fully identify which route the teacher was trained (such as through a BA degree or MA degree). The data used in this study only provided information on the teacher's degree and type of license. Thus, it was possible the teacher received a BA degree outside of education, and then later decided to become a teacher and was trained in a master's degree program. To account for this, I only selected those that had a BA degree *and* had an initial license due to their BA degree. Although this lessened the problem of knowing which program the teacher was trained by, it also reduced the sample of teachers and eliminated all master training programs from the study.

It was also difficult to identify the specific program requirements and elements that each individual candidate was trained with, as there was no way to know whether candidates were exposed to all elements and completed program requirements. For instance, it is possible that candidates substituted requirements, or they took more or less of a requirement (such as math or English courses). One way to address this could be to use candidate transcripts to identify which program they were trained by and which courses candidates took to capture their training more accurately. From there, one could then link transcript data to subsequent outcome data of the candidate/teacher. The limitation of this research design is in obtaining transcripts linked to outcome data for candidates/teachers from multiple programs across the state. However, this is something that could be done on a smaller scale or in partnership with training programs and the state. This approach is further limited in that it does not drill down into specific program elements (namely, training elements that occur within courses) and would only be able to determine whether requirements had a relationship to program outcomes. That said, this approach could serve as a more accurate depiction of candidate training experiences and, from there, researchers could then investigate elements within those requirements that have a significant relationship to student achievement and/or teacher attrition.

Finally, one future improvement in the research design of this study would be to sample all programs within the state of North Carolina. As noted previously, I only surveyed programs that had a significant relationship to student achievement, which was intended to provide a qualitative understanding of effective programs and the elements within those programs. However, as this reduced the total variance in program elements, and subsequently reduced statistical power in the quantitative analysis, a different sampling approach could be explored in future analyses. By including all programs, and not just those that saw a significant relationship, elements could be further differentiated in a quantitative analysis.

6.7 Future Research on Teacher Training

As findings in this study were limited to elements of teacher training that have a significant relationship to student achievement and teacher attrition, even more research questions remain. One looming question is which elements, if any, effect both student achievement and teacher attrition. Programs are often evaluated on both metrics (at least in public rankings), so further investigation into this should be pursued. In this study, these outcomes were explored separately as including student achievement data in the attrition analysis would have greatly reduced the sample size. However, with a larger dataset that captures both outcomes, this would provide more insight into whether any element has a relationship to student achievement and teacher attrition when both are present in the analysis.

One element that was significant in both this study and in Preston's (2017) study on program elements and student achievement in middle and high school was instructional technology. Although this study defined instructional technology differently, both showed significance to student achievement, and it was also significant in the attrition portion of this study. Given these findings, it is important to further investigate this element to examine what specifically matters within this element. For instance, what matters most: general exposure, the method by which candidates are trained in using the technology, or the repeated use of the technology within the student teaching experience? Deep investigation into the relationship to this element and student achievement/attrition may be one method to further illuminate whether it is the training that adds to a teacher's effectiveness, the practice, or a combination. Furthermore, any study that delves into methods used to train candidates (beyond the number of courses that expose them to the technology) should incorporate its relationship to both student achievement and attrition. This would fill a strong void in the current research on instructional technology because although literature on instructional technology and student achievement is growing, there is scant research on its effects on teacher attrition.

In every survey response and interaction with programs in this study, there was reference to the implementation of edTPA. This study did not explore edTPA, as programs in North Carolina are just now finalizing the implementation of this assessment tool (with full implementation by the fall of 2020). However, as this is a substantial endeavor nationally and particularly in the state of

North Carolina, edTPA scores should be considered as a third outcome metric in this analysis. The analysis of program elements and edTPA scores may provide greater information as to which elements are related to a candidate's teaching development. For instance, examining whether candidates from programs that require a higher number of method courses have significantly higher scores in instruction and student engagement tasks. One could then investigate how those findings then translate to student achievement and attrition.

While this study did not include teachers who taught in charter schools, given that charter schools are growing in number throughout the state of North Carolina, it is increasingly important to investigate teacher quality within charter schools. Specifically, as teachers in charter schools are not required to have a license and they are not necessarily trained through traditional training methods, an analysis on the achievement of charter school teachers tied to the training of the teacher, could further shed light on elements of training that have an impact on teacher effectiveness. Perhaps even more important would be to expand this analysis to teacher attrition and training for those within charter school charter schools have significantly higher attrition rates. Which could then serve to inform charter schools as they recruit and work to keep their teachers within their schools as well as the profession (Cano, et al., 2017).

Although the focus of this study was the effects of teacher training on all elementary student populations in North Carolina public schools, I added an investigatory analysis on students from low income families. This initial analysis showed few significant programs (likely due to small sample sizes), but there are some notable changes that would be important to explore further. For instance, some programs show a strong negative effect on student achievement, indicating that there is a difference in the effects of a teacher's preparation for those teaching low income and middle to high incomes students. While I did not investigate these findings further, it does suggest there is much that can be explored between teacher training and effects on low income students. For instance, are there differences in program training techniques between programs that have a positive effect on low-income students compared to those that have no effect or a negative effect? Do the elements that this study found to be positively tied to student achievement gain have the same effect in elementary schools where students are predominately low-income? Furthermore, while school poverty level is not necessarily a predictor of a teacher's attrition, are there training methods that may better prepare teachers of low-income students (and therein lead to lower attrition rates)? Analysis into these questions would provide deeper understanding into the effects of teacher training for different student populations.

One additional analysis that would serve programs and districts alike (as well as the broad literature base) would be to evaluate the cost and benefit of adding certain elements to programs, and subsequently hiring teachers from those programs. In each interview and throughout survey responses, programs noted that while there were elements they would like to include, they were unable to do it because of financial constraints. Furthermore, as the effects shown in this study are both small (specifically in the student achievement analysis) and large (in the attrition) analysis, an assessment of the costs compared to the benefits could prove informative to the community. This is especially true as programs continue to be reviewed, and districts look to programs to train their teachers to be effective for the longevity of their careers.

6.8 Conclusion

I turn once more to my initial research question: what matters in teacher training? The answer is, as most research findings are, complex. For instance, results showed that attendance at certain training programs can increase student achievement in math and reading, as well as a teacher's attrition. Furthermore, there were signs that specific courses in the training program (such as subject matter only) can also impact a student's learning and teacher attrition. However, as I further inspected the classes and training itself, those variables got harder to define and ultimately to model in the analysis. That said, there were two elements of training techniques themselves that showed an impact on student achievement and teacher attrition: recording a candidate's teaching practice in student teaching experiences and lessons on instructional technology (that are embedded in required training courses). Such findings suggest that elements within training likely matter as much as the overall training requirements (if not more), and that more research is needed on the ground and within programs to further test both the effects of elements themselves, as well as the quality of the element on student and teacher outcomes.

Appendix A Interview Protocol

Institution Name:

Interviewee:

Title:

Thank you for your willingness to participate in this study. As mentioned in the human subject consent form you completed prior to this interview, all information discussed today will be confidential and you will have the opportunity to review your responses after the interview is completed. Furthermore, your participation is completely voluntary and you may stop at anytime if you no longer wish to participate. There are no intended benefits or foreseen risks to participating in this interview.

This interview is expected to last no longer than 1 hour, however, there is a possibility I will contact you afterward only for clarification of responses.

You have been selected to speak with me today as you have been identified as someone who has deep knowledge of the elementary training program with which you work. My research study focuses on what elements of teacher training may effect student achievement and/or teacher attrition. As a result, I will be asking questions about elements of your program (such as faculty experience and student teacher placement procedures) that are not publicly available. This study does not aim to evaluate your techniques, offerings, overall program, or university teaching program. Instead, this study wants to better understand the on-the-ground experiences of teacher candidates and the vast program variation offered to elementary teachers in North Carolina. You, your program, and university will never be identified in this research study.

A. Interviewee Background

   a. How long have you been in your present position and at your institution?

   b. Briefly describe your role as it relates to the elementary teacher training program at your institution:

B. Timeline of Training Requirements and Reforms

   a. There are lots of reforms going on today in the teacher preparation space and as such I want to better understand if there are any reform efforts that are present at your program. To get at this information all questions about training offerings will ask when the current processes and procedures were in place. Before we get into those questions however, I would like to ask:

      i. Is your program currently undergoing any reform efforts?

         1. Specifically, what are those efforts aiming to reform?

         2. Can you tell me what those reform look like? Please list specific examples to program changes included in these reforms.

      ii. If the answer above is no, can you tell me when was the most recent program overhaul?

         1. Specifically, what were those efforts aiming to reform?

2. Can you tell me what those reform were? Please list specific examples to program changes included in those reforms.

C. Teacher Training:

a. The next series of questions will ask specifically about requirements and training experiences of your teacher candidates. First I will start with questions regarding faculty qualifications and retention. Questions will then progress to questions on assessment of a candidate's teaching, cooperating/mentoring teacher qualifications and expected feedback, use of data and edTPA scores to guide training and program review, and will end with an opportunity to share elements about your program that are unique and may influence the preparation of your teachers, and to provide feedback on elements that are not otherwise covered but may influence a teacher's preparedness.

D. Faculty

a. This series of questions will ask about faculty qualifications and retention in your elementary teacher training program.

i. What are the qualifications for faculty to teach in your program?

ii. What rewards, if any, do faculty receive for innovative teaching and learning strategies?

1. What are the primary teaching strategies faculty use in their method classes?

iii. What resources are available to faculty to improve their teaching and assessment?

iv. How are faculty teaching practices assessed?

1. Are there peer review practices? Are teaching evaluations primarily based on student evaluations? Is there a department emphasis on faculty teaching?

v. Over the last five years, has your program seen any faculty turnover?

1. Can you describe that turnover? Approximately how many faculty are new or have left your program? Have you been able to fill those positions with equally qualified faculty?

vi. How long have these practices and qualifications been in place?

E. Teacher Candidate Assessment and Feedback

a. How does your program monitor the teaching development of your pre-candidates? Specifically, what tactics are employed to identify strengths and weakness in a candidate's teaching? Similarly, what tactics are employed to address these areas of weaknesses?

b. What are the teaching expectations of your candidates when they graduate and how does your program assess they have met those expectations?

c. Do you use evidence of student learning in these assessments?

i. If so, what evidence is used and how does your program incorporate this into training?

d. How early in a candidate's training do they have teaching experiences? How frequently in their training program do they receive feedback on their training?

e. Can you describe the teaching feedback your teacher candidates receive before they enter their student teaching experience?
    i. Do you use video or other technology to review a candidate's teaching and provide feedback?
    ii. If yes, how do you use video or other technology in program feedback?
f. How long have these practices been in place in your teaching program?

F. Cooperating/Mentoring Teacher Qualifications, Assignment, and Feedback

a. Describe the placement process of teacher candidates to their cooperating or mentoring teacher?
b. What are the program's requirements for a cooperating teacher to be placed with a teacher candidate?
    i. What are the cooperating teacher's required qualifications?
c. What is the required feedback that a cooperating teacher must provide to the teacher candidate?
d. What is the program's role in ensuring a candidate receives this feedback during the student teaching experience?
e. How long have these practices been in place in your teaching program?

G. Use of Data to Inform Teacher Training Such as edTPA

a. Does your program use edTPA assessments in your training?
    i. If yes, how does your program use this?

        ii. What feedback or additional training do you provide to the candidate as a result of their assessment results?

   b. Are there other assessments your program uses to assess a teacher's preparedness that are not previously mentioned?
   c. How long have these practices been in place in your teaching program?

H. Other Efforts:

   a. Are there other areas that are unique to your training program that have not been covered by prior questions? Please provide specific examples and descriptions of these efforts.
   b. What about these efforts are unique to your program in particular?
   c. How long have these efforts been in place in your program?

I. Questions Not Asked:

   a. One goal of this interview is to learn more about the nuances of elementary teacher training programs in North Carolina to build a survey and ask all training programs throughout the state about their program offerings. As a result, I was wondering if there are questions you think would be beneficial in asking of other programs that I have not already asked?
        i. Can you tell me what you think are some of the broad trends you are seeing in North Carolina training programs and the direction programs are taking with their training?

Appendix B Survey Instrument

Survey of Teacher Training Practices

1. Name of the institution you are affiliated with (for the purposes of this survey only):
2. **If you are affiliated/more familiar with a specific teacher training program at your institution, please identify that program below (please select all that apply).**

    Undergraduate Elementary Education

    Undergraduate Elementary Math Education

    Undergraduate Elementary Reading Education

    MAT in Elementary Math Education

    MAT in Elementary Reading Education

    MAT in Elementary Education

    Familiar with all elementary education programs

    Other (please list)

3. **Please indicate approximately how many years you have worked at your university:**

    Less than 1 year

    1-2 years

    3-4 years

    5 or more

4. **How many years have you worked in your current position?**

Less than 1 year

1-2 years

3-4 years

5 or more years

5. **Are you involved in program requirement decisions (such as setting curriculum or program policies) as it relates to the elementary teacher training program at your institution?**

   Yes

   No

Start of Block 1: Faculty Qualifications and Teaching Requirements

6. **Which course levels do non-tenure track faculty teach? Please select all that apply:**

   Introductory education courses

   Content courses (such as foundations in mathematics or reading)

   Pedagogy courses (such as foundations in teaching methodology

   Content Pedagogy (such as teaching methods for mathematics)

   Seminar/student teaching

   All of the above

   Non-tenure track faculty do not teach pre-service candidates

7. **Since 2014, have these faculty teaching assignments changed?**

   If yes, please explain

   No

8. **Are there incentives for faculty to receive professional development of their own teaching practices (for either tenure-track or non-tenure track)? Please select all that apply.**

    University or department financial awards for faculty teaching

    Workshops on faculty teaching techniques sponsored by the program/department/university

    Faculty networks or meetings to discuss teaching strategies (within the program/department—not including professional associations)

    Financial support to attend professional associations specifically aimed at improvement of faculty teaching

    Other (please explain)

9. **Do faculty reviews in your program/department include assessment of faculty teaching practices?**

    Yes

    No

10. **If assessment of faculty teaching is included as part of faculty reviews, please explain how this is used.**

Start of Block 2: In-Service Teaching

11. Please list the course name/number in the curriculum when a candidate has their first in-classroom teaching experience (experience can be informal, teaching one lesson, co-teaching, or other)?

12. When in the candidate's course sequence does the candidate usually have their first in-classroom experience? Please select only one option below.

First Year

Second Year

Third Year

Fourth Year/During Seminar

13. Approximately how many courses within a teacher candidate's training require active teaching of the candidate?

Less than 3

Between 3-5

Between 5-10

Between 10-15

15 or more

14. Do courses in your program have opportunities for students to use computer adaptive teaching experiences (where the candidate can model teaching through an interactive computer program)

Yes, it is required

Yes, but it is optional

No

15. **If the use of computer adaptive teaching experiences is required or encouraged, when in the candidate's course sequence is this offered? Please select all that apply.**

All method courses

Seminar only

All program required

Introductory courses

Optional resource available, not in a specific course

Other course (please list) ____

16. **Are there courses within your program that teach candidates about technology use in their classroom?**

Yes

No

17. **If there are courses within your program that teach candidates about technology use in the classroom, please select all courses that apply below.**

All method courses

Seminar only

Introductory courses

All program required courses except foundational content courses

Other course (list course title)

18. **Approximately how many courses within a teacher candidate's training include lessons on using technology in the classroom?**

    Less than 3

    Between 3-5

    Between 5-10

    Between 10-15

    15 or more

19. **Do courses in your program require candidates to videotape examples of their teaching?**

    Yes, required

    No, but is an option

    No

20. **If videotaping candidate teaching is required (or an option in some courses) which courses include this requirement or option? Please select all that apply.**

    All method courses

    Seminar only

    Introductory courses

    All program required courses except foundational content courses

    Specific course (please list) ______________________________

    Not associated with a specific course

Start of Block 3: Teacher Candidate Assessment and Feedback

21. **When does your program assess the teaching development of your pre-candidates? Please select only one option.**

    Once a year with the candidate's advisor

    At the end of junior year before entering seminar

    Through courses and seminar--no separate assessment

    No overall assessments are present outside of those within specific method courses

22. **Is there a standardized review process of a candidate's teaching before they enter their student teaching experience and/or seminar?**

    Yes

    No

23. **If yes, there is a standardized review process, how is this accomplished? Please select all that apply.**

    Standardized rubric

    Reflective discussion between professor and candidate

    Formal in-person review by advisor (using standardized rubric)

    Part of a specific course (please indicate course and content of course)

    Other (please explain) ____

24. **Does your training program use evidence of elementary student learning in candidate teaching assessments (evidence can include pre-post test scores, overall grades of students in their classrooms, or other rubrics)?**

    Yes

    Yes, but is not required of the program

    No

25. **If examples of student learning are included in candidate teaching reviews, what evidence/method is used? Please select all that apply.**

    Pre-post test scores

    Grades of students in the candidate's classroom

    Reviews of the coordinating teacher

    Other (please provide a brief explanation)

26. **Does your program use mentors for candidates (other than the candidate's cooperating teacher) during in-service training? (These can include emeritus, adjunct, faculty, or other teaching or administrative representatives).**

    Yes

    No

27. **If your program does have mentors (outside of cooperating teachers) please indicate type of mentor. Please select all that apply:**

Emeritus faculty

Adjunct or non-tenure track faculty

Tenured or tenure track faculty

Local teachers

Other (please explain)

28. Is assessment of a candidate's teaching a required part of this mentorship?

Yes

No

29. If candidate teaching assessment is a required part of this mentorship, is this a formal/standardized process or is it an informal assessment?

Formal/Standardized Process

Informal Assessment Process

30. Has assessment of a candidate's teaching changed since 2014?

Yes

No

31. If candidate teaching experience has changed, please indicate which areas below. Please select all that apply.

Use of video technology

Use of student learning in candidate assessment

Frequency of candidate teaching assessment

Candidate teaching assessment before seminar experience

Other (please explain)

Start of Block 4: Qualities Unique to Your Program and Institution

32. **What do students/alumni from your program cite as the most useful part of your program?**

33. **What elements of your program do faculty cite that the program does particularly well?**

34. **Are there other areas that are unique to your training program that you would like to share? Please provide information on these efforts.**

Start of Block 5: Current Program Reforms

35. **Since 2014, has your program undergone any structural reforms (such as course requirements, in-service training hours, different candidate teaching experiences, and/or piloting or use of edTPA)?**

    Yes

    No

36. **Has your program reformed or modified course requirements after 2014?**

    Yes

    No

37. In what year (approximately) were candidate courses reformed, modified, or course requirements changed?

2014

2015

2016

2017

2018

38. If known, please indicate which courses were modified (or added/removed)

39. Has your program changed the number of in-service teaching hours required for candidates after 2014?

Yes

No

40. In what year (approximately) were hours of in-service training reformed, modified, or otherwise changed?

2014

2015

2016

2017

2018

41. If known, please indicate the change in hours

42. Has your program reformed or modified required candidate teaching experiences after 2014?

Yes

No

43. In what year (approximately) were candidate teaching experiences reformed, modified, or otherwise changed?

2014

2015

2016

2017

2018

44. Please indicate the change in candidate teaching experiences (please select all that apply):

Small opportunities throughout the length of the program

Co-teaching opportunities with mentors

Virtual or computer modeled experiences

Other (please explain) ___

45. Has your program added the edTPA assessment to your training requirements?

Yes, it is a requirement for all elementary candidates

Yes, your program is currently piloting edTPA with elementary teaching candidates

No, your program has not used it with elementary teaching candidates

46. If yes, what year was edTPA fully implemented

2014

2015

2016

2017

2018

47. What year did your program begin piloting edTPA with candidates?

2014

2015

2016

2017

2018

48. For which candidates has edTPA been implemented/piloted?

All elementary teaching candidates

Math-only candidates

Reading-only candidates

All teacher education candidates in the School of Education

Other (please explain)

49. If your program has implemented (or piloted) edTPA, how has will this be used in your elementary training program (select all that apply)

Program assessment and review

Candidate assessment

Informational only—program or candidate assessment will come from other evaluative measures

Undetermined

Other (please explain)

Start of Block: Thank you

Thank you for completing this survey. Please fill out the below information to receive your $25 incentive.

50. Name

51. Physical mailing address to send your $25 incentive

Appendix C Additional Analysis of High Poverty Schools

Table 6.1. Reading Achievement Gains at High Poverty Schools by Training University

	Multilevel Model Model 6		School Fixed Effects Model Model 7	
	β	SE	β	SE
Intercept	3.26**	1.68	3.06**	1.46
Student Level				
Year	0.42***	0.13	0.35*	0.19
Student Poverty	-0.32	0.23	-0.33*	0.19
Minority Student	-0.22	0.26	-0.23	0.22
Male Student	0.28*	0.19	0.28	0.20
ESL Student	0.81**	0.29	0.81**	0.28
Teacher Level				
%Class_Minority	-1.41**	0.73	-0.79	1.15
%Class_Poverty	1.29	1.02	1.10	1.40
%Class_ESL	0.10	0.87	0.24	1.19
Minority Teacher	0.19	0.36	0.38	0.53
Male Teacher	-0.06	0.48	0.23	0.61
Master's Degree	-0.14	0.49	-0.57	0.55
1st Year Teacher	-0.70*	0.46	-0.33	0.60
2nd Year Teacher	-0.63	0.46	-0.26	0.51
3rd Year Teacher	-0.59	0.42	-0.62	0.64
4th Year Teacher	-0.63*	0.40	-0.33	0.61
University 25	4.31**	1.62	4.16***	1.26
University 13	3.30**	1.56	3.42*	1.78
University 18	3.03*	1.60	2.54*	1.41
University 14	2.95*	1.72	2.47*	1.59
University 15	2.95*	1.60	2.48*	1.45
University 9	2.79*	1.68	3.32**	1.49
University 23	1.96	1.57	0.88	1.54
University 30a	1.96	1.54	3.56**	1.62
University 20	1.95	1.60	1.87	2.01
University 16	1.79	1.61	0.69	1.46
University 2	1.75	1.43	1.44	1.39
University 3	1.74	1.48	1.25	1.54
University 5	1.65	1.48	1.68	1.39
University 1	1.63	1.44	1.42	1.41
University 10	1.60	1.47	2.18	1.54
University 4	1.50	1.47	1.65	1.30
University 6	1.42	1.53	1.58	1.41
University 22	1.36	1.70	0.21	1.45

Table 6.1 Continued

	Multilevel Model Model 6		School Fixed Effects Model Model 7	
	β	SE	β	SE
University 11	1.19	1.44	3.76**	1.34
University 27	1.09	1.67	1.09	1.81
University 17	1.02	1.61	0.47	1.83
University 8	0.89	1.58	0.58	1.64
University 21	0.87	1.56	-0.01	1.69
University 12	0.81	1.61	-0.01	1.71
University 19	0.73	2.19	-0.32	1.86
University 24	0.72	1.65	0.09	1.24
University 26	0.60	1.41	3.63**	1.31
University 7	0.47	1.56	1.08	1.67
University 28	-2.19	2.38	-6.96***	1.76
University 29	-	-	-	-
Student Variance[39]	35.91		38.07	
Student Variance Explained	0.00		0.01	
Teacher Variance	2.82		11.36	
Teacher Variance Explained	0.11		-0.12	
ICC	0.06		0.23	
Total Variance Explained	0.01		0.02	
Number of Students	4,033		4,033	
Number of Teachers	469		469	

***p< 0.001 level. **p< 0.05 level * p<0.1 level

[39] Variance for the multilevel baseline model is: 36.07(student variance), 3.18 (teacher variance), and 0.08 (ICC). Variance for the school fixed effect baseline model is: 38.44 (student variance), 10.18 (teacher variance), and 0.21 (ICC).

Table 6.2. Math Achievement Gains at High Poverty Schools by Training University

	Multilevel Model Model 6		School Fixed Effects Model 7	
	β	SE	β	SE
Intercept	3.11*	1.67	3.05*	1.81
Student Level				
Year	-0.62***	0.12	-0.53***	0.14
Student Poverty	-0.21	0.20	-0.26	0.21
Minority Student	0.14	0.24	0.18	0.28
Male Student	-0.08	0.17	-0.09	0.18
ESL Student	1.00***	0.24	0.93***	0.27
Teacher Level				
%Class_Minority	-1.85**	0.75	-1.51	1.23
%Class_Poverty	0.79	1.01	1.26	0.97
%Class_ESL	-1.77*	0.93	-1.91*	1.25
Minority Teacher	0.39	0.36	0.79	0.65
Male Teacher	-0.19	0.35	0.30	0.40
Master's Degree	-0.69*	0.44	-0.52	0.46
1st Year Teacher	-1.47**	0.48	-1.11**	0.42
2nd Year Teacher	-0.72*	0.45	-0.47	0.42
3rd Year Teacher	0.25	0.43	0.61	0.48
4th Year Teacher	0.47	0.42	0.65*	0.38
University 19	2.78	1.58	2.05	1.55
University 9	2.12	1.48	0.94	1.36
University 6	1.91	1.61	0.67	1.44
University 3	1.34	1.46	-0.06	1.41
University 13	1.33	1.66	-0.11	1.83
University 10	1.27	1.56	-0.99	1.56
University 23	1.18	1.71	-0.98	1.94
University 21	1.06	1.66	-0.61	1.60
University 11	1.02	2.06	-0.11	3.02
University 27	0.98	2.13	1.89	2.48
University 25	0.82	1.52	-0.48	1.45
University 20	0.77	1.75	0.43	1.97
University 24	0.76	2.26	-0.48	2.09
University 17	0.55	1.63	-0.82	1.89
University 1	0.53	1.42	0.71	1.21
University 15	0.41	1.52	-1.39	1.25
University 4	0.36	1.43	-0.79	1.27
University 26	0.34	1.71	-3.66**	1.41
University 5	0.05	1.46	-1.25	1.32
University 22	-0.17	1.72	-0.78	1.52
University 14	-0.21	1.76	-0.80	1.82
University 30b	-0.33	1.48	-1.20	1.45

Table 6.2 Continued

	Multilevel Model Model 6		School Fixed Effects Model 7	
	β	SE	β	SE
University 12	-0.45	1.57	-1.61	1.43
University 28	-0.73	2.22	2.45	1.82
University 2	-0.77	1.43	-0.06	1.38
University 18	-1.22	1.64	-2.52*	1.56
University 7	-1.58	1.64	-2.03	1.48
University 8	-1.58	1.57	-2.20	1.52
University 16	-1.88	1.66	-3.74*	1.48
University 29	-2.88**	1.45	-1.64	1.48
Student Variance	28.43		31.92	
Student Variance Explained[40]	0.01		0.03	
Teacher Variance	5.17		9.42	
Teacher Variance Explained	0.18		0.02	
ICC	0.15		0.23	
Total Variance Explained	0.04		0.03	
Number of Students	4,686		4,686	
Number of Teachers	499		499	

***p< 0.001 level. **p< 0.05 level * p<0.1 level

[40] Variance for the multilevel baseline model is: 28.80 (student variance), 6.29 (teacher variance), and 0.18 (ICC). Variance for the school fixed effect baseline model is: 32.83 (student variance), 9.61 (teacher variance), and 0.23 (ICC).

Bibliography

AACTE (2018). Colleges of Education: A National Portrait. Washington DC. Retrieved from: https://secure.aacte.org/apps/rl/res_get.php?fid=4178&ref=rl

AACTE (2019). State Participation in edTPA. Retrieved 2/1/19 from: http://edtpa.aacte.org/state-policy

Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and student achievement in the Chicago public high schools. *Journal of labor Economics*, *25*(1), 95-135.

Adnot, M., Dee, T., Katz, V., & Wyckoff, J. (2017). Teacher turnover, teacher quality, and student achievement in DCPS. *Educational Evaluation and Policy Analysis*, *39*(1), 54-76.

Aldeman, C., Carey, K., Dillon, E., Miller, B., & Silva, E. (2011). A measured approach to improving teacher preparation. *Education Sector Policy Brief*, 5-22.

Al Otaiba, S., Lake, V. E., Greulich, L., Folsom, J. S., & Guidry, L. (2012). Preparing beginning reading teachers: An experimental comparison of initial early literacy field experiences. *Reading and Writing, 25*(1), 109–129.

Allen, M. (2003). Eight questions on teacher preparation: What does the research say? A summary of findings. Denver, CO: Education Commission of the States.

Allison, P. D. (2005). Fixed effects regression methods for longitudinal data using SAS. Sas Institute.

Allison, P. D. (2009). *Fixed effects regression models* (Vol. 160). SAGE publications.

Ball, D.L., Thames, M.H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? Journal of teacher education, 59(5), 389-407.

Barnes, G., Crowe, E., & Schaefer, B. (2007). The Cost of Teacher Turnover in Five School Districts: A Pilot Study. National Commission on Teaching and America's Future.

Bastian, K.C., & Henry, G.T. (2015). Teachers Without Borders: Consequences of Teacher Labor Force Mobility. Educational Evaluation and Policy Analysis 37(2): 163-183.

Bastian K.C., Henry, G.T., Pan, Y., & Lys, D. (2016). Teacher candidate performance assessments: Local scoring and implications for teacher preparation program improvement. Teaching and Teacher Education, 59, 1-12.

Bastian, K. C., Patterson, K. M., & Pan, Y. (2018). Evaluating teacher preparation programs with teacher evaluation ratings: Implications for program accountability and improvement. *Journal of Teacher Education, 69*(5), 429-447.

Belson, S. I., & Husted, T. A. (2015). Impact of National Board for the Professional Teaching Standards (NBPTS) certification on student achievement. *education policy analysis archives*, *23*, 91.

Benedict, A., Holdheide, L., Brownell, M., & Marshall Foley, A. (2016). Special Issues Brief—Learning to Teach: Practice-Based Preparation in Teacher Education. *Center on Great Teachers & Leaders. American Institutes for Research.* Washington D.C.

Boonen, T., Van Damme, J., & Onghena, P. (2014). Teacher effects on student achievement in first grade: which aspects matter most? *School Effectiveness and* School Improvement, 25(1), 126-152.

Boyd, Grossman, Lankford, Loeb, & Wyckoff (2006) How changes in entry requirements alter the teacher workforce and affect student achievement**.** *Education Finance and Policy*, 1 (2),176-216

Boyd, D. J., Grossman, P. L., Lankford, H., Loeb, S., & Wyckoff, J. (2009). Teacher preparation and student achievement. *Educational Evaluation and Policy Analysis, 31*(4), 416-440.

Boyd, D., Lankford, H., Loeb, S., Rockoff, J., & Wyckoff, J. (2008). The narrowing gap in New York City teacher qualifications and its implications for student achievement in high-poverty schools. *Journal of Policy Analysis and Management: The Journal of the Association for Public Policy Analysis and Management*, *27*(4), 793-818.

Boyd, D., Lankford, H., Loeb, S., Ronfeldt, & Wyckoff, J. (2011). The role of teacher quality in retention and hiring: Using applications to transfer to uncover preferences of teachers and schools. Journal of Policy Analysis and Management, 30(1), 88-110.

Buddin, R., & Zamarro, G. (2009). Teacher qualifications and student achievement in urban elementary schools. *Journal of Urban Economics*, *66*(2), 103-115.

Braga, M., Paccagnella, M., & Pellizzari, M. (2014). Evaluating students' evaluations of professors. Economics of Education Review, 41-71-88.

Bryk, A. S., & Raudenbush, S. W. (1992). Hierarchical linear models for social and behavioral research: Applications and data analysis methods.

Bryan, B.A., Hougen, M., & Nelson, K. (2018). Leading on Literacy: Challenges and Opportunities in Teacher Preparation Across the University of North Carolina System. *The University of North Carolina System.* Chapel Hill, North Carolina.

Campbell, P. F., Nishio, M., Smith, T. M., Clark, L. M., Conant, D. L., Rust, A. H., & Choi, Y. (2014). The relationship between teachers' mathematical content and pedagogical knowledge, teachers' perceptions, and student achievement. *Journal for Research in Mathematics Education*, *45*(4), 419-459.

Cameron, A.C., & Miller, D.L. (2015). A practitioner's guide to cluster-robust inference. Journal of human resources, 50(2), 317-372.

Cano, S.L., Flores, B.B., Claeys, L., & Sass, D.A. (2017). Consequences of Educator Stress on Turnover: The Case of Charter Schools. In *Educator Stress* (pp. 127-155). Springer, Cham.

Carruthers, C.K. (2012). New schools, new students, new teachers: Evaluating the effectiveness of charter schools. *Economics of Education Review*, 31(2), 280-292.

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., & Yagan, D. (2011). How does your kindergarten classroom affect your earnings? Evidence from Project STAR. *The Quarterly Journal of Economics*, *126*(4), 1593-1660.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2011). *The long-term impacts of teachers:* Teacher value-added and student outcomes in adulthood (No. w17699). National Bureau of Economic Research.

Chingos, M. M., & Peterson, P. E. (2011). It's easier to pick a good teacher than to train one: Familiar and new results on the correlates of teacher effectiveness. *Economics of Education Review*, *30*(3), 449-465.

Chingos, M.M. (2014). Who Profits from the Master's Degree Pay Bump for Teachers? Brookings Institute. Retrieved December 20, 2018 from: https://www.brookings.edu/research/who-profits-from-the-masters-degree-pay-bump-for-teachers/

Chung, R. R. (2008). Beyond assessment: Performance assessments in teacher education. Teacher Education Quarterly, 35(1), 7-28.

Clark, D., Martorell, P., & Rockoff, J. (2009). School Principals and School Performance. Working Paper 38. *National Center for Analysis of longitudinal data in Education research*. Accessed December 12, 2018 from: https://files.eric.ed.gov/fulltext/ED509693.pdf

Clark, M. A., Chiang, H. S., Silva, T., McConnell, S., Sonnenfeld, K., Erbe, A., & Puma, M. (2013). The effectiveness of secondary math teachers from Teach For America and the Teaching Fellows programs. *Princeton, NJ: Mathematica Policy Research*. Retrieved December 14, 2018 from: https://files.eric.ed.gov/fulltext/ED544171.pdf

Clark, M. A., Isenberg, E., Liu, A. Y., Makowsky, L., & Zukiewicz, M. (2015). Impacts of the Teach For America investing in innovation scale-up. *Princeton, NJ: Mathematica Policy Research*. Retrieved December 14, 2018 from: https://www.mathematica-mpr.com/our-publications-and-findings/publications/impacts-of-the-teach-for-america-investing-in-innovation-scaleup

Close, K., Amrein-Beardsley, A., & Collins, C. (2018). State-Level Assessment and Teacher Evaluation Systems After the Passage of the Every Student Succeeds Act: Some Steps in the Right Direction. National Education Policy Center.

Clotfelter, C., Glennie, E., Ladd, H., & Vigdor, J. (2008). Would higher salaries keep teachers in high-poverty schools? Evidence from a policy intervention in North Carolina. Journal of Public Economics, 92(5-6), 1352-1370.

Clotfelter, C.T., Hemelt, S.W., Ladd, H.F., & Turaeva, M.(2019). School Segregation in the era of immigration, school choice and color-blind jurisprudence—the case of North Carolina. (EdWorkingPaper: 19-101).

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2007). Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. *Economics of Education Review*, *26*(6), 673-682.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2010). Teacher credentials and student achievement in high school a cross-subject analysis with student fixed effects. *Journal of Human Resources*, *45*(3), 655-681.

Clotfelter, C.T., Ladd, H.F., Vigdor, J.L., & Wheeler, J. (2007). High-poverty schools and the distribution of teachers and principals (CALDER Working Paper 1). Washington, DC: Urban Institute

Coenen, J., Cornelisz, I., Groot, W., Maassen van den Brink, H., & Van Klaveren, C. (2018). Teacher Characteristics and Their Effects On Student Test Scores: A Systematic Review. *Journal of Economic Surveys*, 32 (3), 848-877.

Coggshall, J. G., Bivona, L., & Reschly, D. J. (2012). Evaluating the Effectiveness of Teacher Preparation Programs for Support and Accountability. Research & Policy Brief. *National Comprehensive Center for Teacher Quality*.

Cohen, J., Ruzek, E., & Samdilos, L. (2018). Does teaching quality cross subjects? Exploring consistency in elementary teacher practice across subjects. AERA Open, 4(3).

Condon W., Iverson, E.R., Manduca,C.A., Rutz, C., & Willett, G. (2016),. Faculty Development and Assessing the Connections. Indiana University Press.

Council of Chief State School Officers (2016). Accountability in Teacher Preparation: Policies and Data in the 50 States & DC. Washington, D.C. Retrieved from: https://ccsso.org/sites/default/files/2017-10/50StateScan092216.pdf

Council of Chief State School Officers (2017). Transforming Educator Preparation: Lesson Learned from Leading States. Washington, D.C. Retrieved from: https://www.ccsso.org/sites/default/files/2017-11/CCSSO%20Educator%20Preparation%20Playbook.pdf

Cowan, J., & Goldhaber, D. (2016). National board certification and teacher Effectiveness: Evidence from Washington State. *Journal of Research on Educational Effectiveness*, *9*(3), 233-258.

Croninger, R. G., Rice, J. K., Rathbun, A., & Nishio, M. (2007). Teacher qualifications and early learning: Effects of certification, degree, and experience on first-grade student achievement. *Economics of Education Review*, *26*(3), 312-324.

Crowe, E. (2011). Race to the Top and Teacher Preparation: Analyzing State Strategies for Ensuring Real Accountability and Fostering Program Innovation. Center for American Progress. Washington D.C. Retrieved from: http://www.americanprogress.org/issues/2011/03/pdf/teacher_preparation.pdf

Darling-Hammond, L. (2015). Can value added add value to teacher evaluation?. *Educational Researcher*, *44*(2), 132-137.

Darling-Hammond, L. (2016). Research on teaching and teacher education and its influences on policy and practice. *Educational Researcher*, *45*(2), 83-91.

Darling-Hammond, L., Berry, B., & Thoreson, A. (2001). Does teacher certification matter? Evaluating the evidence. Educational evaluation and policy analysis, 23(1), 57-77.

Darling-Hammond, L. & Carver-Thomas, D. (2017). Teacher turnover: Why it matters and what we can do about it. Washington, DC: Learning Policy Institute.

Darling-Hammond, L., Newton, X., & Wei, R. C. (2010). Evaluating teacher education outcomes: A study of the Stanford Teacher Education Programme. *Journal of Education for Teaching*, *36*(4), 369-388.

Dieker, L. A., Straub, C., Hughes, C. Hynes, M. C., & Hardin, S. E. (2014). Learning from virtual students. *Educational Leadership, 71*(8), 54–58.

Engle, U., Jann, B., Lynn, P., Scherpenzeel, A., & Sturgis, P. (Eds.). (2014). Improving survey methods: Lessons from recent research. Routledge.

Epple, D., Romano, R., & Zimmer, R. (2016). Charter Schools: A survey of research on tgeir characteristics and effectiveness. In *Handbook of the Economics of Education* (Vol. 5, 139-208). Elsevier.

Fahle, E., Reardon, S., Kalogrides, D., Weathers, E., & Jang, H. (2020). Racial Segregation and School Poverty in the United States, 1999-2016. Race and Social Problems, 12(1), 42-56.

Feuer, M. J., Floden, R. E., Chudowsky, N., and Ahn, J. (2013). *Evaluation of teacher* preparation programs: Purposes, methods, and policy options. Washington, DC: National Academy of Education.

Friedrichsen, P. J., Abell, S. K., Pareja, E. M., Brown, P. L., Lankford, D. M., & Volkmann, M. J. (2009). Does teaching experience matter? Examining biology teachers' prior knowledge for teaching in an alternative certification program. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, *46*(4), 357-383.

Gansle, K. A.; Noell, G.H. & Burns, J.M. (2012). Do Student Achievement Outcomes Differ Across Teacher Preparation Programs? An Analysis of Teacher Education in Louisiana. Journal of Teacher Education. 63(5).

Garet, M. S., Heppen, J. B., Walters, K., Parkinson, J., Smith, T. M., Song, M., & Borman, G. D. (2016). Focusing on Mathematical Knowledge: The Impact of Content Intensive Teacher Professional Development. NCEE 2016-4010. *National Center for Education Evaluation and Regional Assistance*.

Garrett, R., & Steinberg, M. P. (2015). Examining teacher effectiveness using classroom observation scores: Evidence from the randomization of teachers to students. *Educational Evaluation and Policy Analysis*, *37*(2), 224-242.

Garson, G. (2020). Multilevel Modeling: Applications in STATA, IBM, SAS, R, & HLM. SAGE Publications.

Gitomer, D., Martinez, J., Battey, D., Hyland, N.E. (2019). Assessing the Assessment: Evidence of Reliability and Validity in the edTPA. American Educational Research Journal.

Glazerman, S., Mayer, D., & Decker, P. (2006). Alternative routes to teaching: The impacts of Teach for America on student achievement and other outcomes. Journal of Policy Analysis and Management: The Journal of the Association for Public Policy Analysis and Management, 25(1), 75-96.

Goff, P., Yang, H., Yang, M., Batt, L., Xie, X., You, E., & Braaten, M. (2020). A New Evaluation Approach for Teacher Preparation Programs Using Labor Market Competitiveness of Teacher Applicants. WCER Working Paper NO. 2020-3.

Goldhaber, D. (2007). Everyone's doing it, but what does teacher testing tell us about teacher effectiveness?. *Journal of human Resources*, *42*(4), 765-794.

Goldhaber, D. (2019). Evidence-based teacher-preparation: Policy context and what we know. Journal of Teacher Education, 70(2), 90-101.

Goldhaber, D., & Anthony, E. (2007). Can teacher quality be effectively assessed? National board certification as a signal of effective teaching. *The Review of Economics and Statistics*, *89*(1), 134-150.

Goldhaber, D. & Brewer, D. J. (2000). Does teacher certification matter? High school teacher certification status and student achievement. *Educational evaluation and policy analysis*, *22*(2), 129-145.

Goldhaber, D., & Cowan, J. (2014). Excavating the Teacher Pipeline: Teacher Preparation Programs and Teacher Attrition. *Journal of Teacher Education*, 65(5), 449-462.

Goldhaber, D., Gross, B, & Player, D. (2011). Teacher career paths, teacher quality, and persistence in the classroom: Are public schools keeping their best? *Journal of Policy Analysis and Management*, 30(1), 57-87.

Goldhaber, D., Krieg, J. M., & Theobald, R. (2016). Does the match matter? Exploring whether student teaching experiences affect teacher effectiveness and attrition. National Center for Analysis of Longitudinal Data in Education Research Working Paper No, 149.

Goldhaber, D., & Liddle, S. (2011). The Gateway to the Profession. Assessing Teacher Preparation Programs Based on Student Achievement. CEDR Working Paper 2011-2. University of Washington, Seattle, WA.

Goldhaber, D., Liddle, S., & Theobald, R. (2013). The gateway to the profession: Assessing teacher preparation programs based on student achievement. *Economics of Education Review*, *34*, 29-44.

Goldhaber, D., & Ronfeldt, M. (2020). Toward causal evidence on effective teacher preparation. Linking Teacher Preparation Program Design and Implementation to Outcomes for Teachers and Students, 211.

Gray, L., & Taie, S. (2015). Public School Teacher Attrition and Mobility in the First Five Years: Results from the First through Fifth Waves of the 2007-2008 Beginning Teacher Longitudinal Study. First Look. NCES 2015=337. National center for education statistics.

Großschedl, J., Mahler, D., Kleickmann, T., & Harms, U. (2014). Content-related knowledge of biology teachers from secondary schools: Structure and learning opportunities. *International Journal of Science Education*, *36*(14), 2335-2366.

Gronqvist, E., & Vlachos, J. (2008). One size fits all? The effects of teacher cognitive and non-cognitive abilities on student achievement (No. 2008:25). Institute for Labour Market Policy Evaluation. Accessed December 12, 2018 from: https://www.econstor.eu/bitstream/10419/45742/1/587127066.pdf

Hanushek, E. A. (1992). The trade-off between child quantity and quality. *Journal of* political economy, 100(1), 84-117.

Hanushek, E. A. (2011). The economic value of higher teacher quality. *Economics of* Education Review, 30(3), 466-479.

Hanushek, E. A., Piopiunik, M., & Wiederhold, S. (2014). *The value of smarter teachers: International evidence on teacher cognitive skills and student performance* (No. w20727). National Bureau of Economic Research. Accessed December 12, 2018 from: https://www.nber.org/papers/w20727.pdf

Hanushek, E. A., & Rivkin, S. G. (2006). Teacher Quality. Hanushek, E.A., and Welch F. (Eds.). Handbook of Economics of Education. 2 (18). Pgs. 1051-1075.

Hanushek, E. A., & Rivkin, S. G. (2010). Generalizations about using value-added measures of teacher quality. *American Economic Review*, *100*(2), 267-71.

Harris, D.N., Ingle, W.K., & Rutledge, S.A. (2014). How teacher evaluation methods matter for accountability: A comparative analysis of teacher effectiveness ratings by principals and teacher value-added measures. America Educational Research Journal, 51(1), 73-112.

Harris, D.N., & Sass, T.R. (2007). Teacher Training, Teacher Quality, and Student Achievement. CALDER Working Paper 3. Accessed July 10, 2015 from: http://files.eric.ed.gov/fulltext/ED509656.pdf

Harris, D. N., & Sass, T. R. (2009). The effects of NBPTS-certified teachers on student achievement. Journal of Policy Analysis and Management: The Journal of the Association for Public Policy Analysis and Management, 28(1), 55-80.

Harris, D. N., & Sass, T. R. (2011). Teacher training, teacher quality and student achievement. *Journal of public economics*, *95*(7-8), 798-812.

Henry, G. T., Bastian, K. C., & Fortner, C. K. (2011). Stayers and leavers: Early-career teacher effectiveness and attrition. *Educational Researcher*, *40*(6), 271-280.

Henry, G. T., Bastian, K. C., & Smith, A. A. (2012). Scholarships to recruit the ‘‘best and brightest’’ into teaching: Who is recruited, where do they teach, how effective are they, and how long do they stay? Educational Researcher, 41(3), 83–92.

Henry, G.T., Bastian, K.C., Fortner, C.K., Kershaw, D.C., Purtell, K.M., Thompson, C.L. & Zulli, R.A. (2014). Teacher Preparation Policies and Their Effects on Student Achievement. Education Finance and Policy. 9 (3). Pgs. 264-303.

Henry, G. T., Campbell, S. L., Thompson, C. L., Patriarca, L. A., Luterbach, K. J., Lys, D. B., & Covington, V. M. (2013). The predictive validity of measures of teacher candidate programs and performance: Toward an evidence-based approach to teacher preparation. *Journal of Teacher Education*, *64*(5), 439-453.

Henry, G.T., Kershaw, D.C., Zulli, R.A., Smith, A.A. (2012). Incorporating Teacher Effectiveness Into Teacher Preparation Program Evaluation. Journal of Teacher Education. 63(5).

Henry, G.T., Patterson, K.M., Campbell, S.L., & Yi, P. (2013). UNC Teacher Quality Research: 2013 Teacher Preparation Program Effectiveness Report. Education Policy Initiative at Carolina. University of North Carolina at Chapel Hill.

Henry, G. T., Purtell, K. M., Bastian, K. C., Fortner, C. K., Thompson, C. L., Campbell, S. L., & Patterson, K. M. (2014). The effects of teacher entry portals on student achievement. *Journal of Teacher Education*, *65*(1), 7-23.

Henry, G. T., Thompson, C. L., Fortner, C. K., Zulli, R. A., & Kershaw, D. C. (2010). The impact of teacher preparation on student learning in North Carolina public schools. *Carolina Institute for Public Policy*: University of North Carolina—Chapel Hill. Retrieved from: http://eqdashboard.northcarolina.edu/wp-content/uploads/2015/11/Impact-of-Educator-Preparation1.pdf

Higher Education Opportunity Act (2008). 20 U.S.C §§ 1021-1041.

Hill, H.C., Ball, D.L., & Schilling, S.G. (2008). Unpacking pedagogical content knowledge: Conceptualizing and measuring teachers' topic-specific knowledge of students. Journal for research in mathematics education 372-400.

Hill, H. C., Kapitula, L.,& Umland, K. (2011). A validity argument approach to evaluating teacher value-added scores. *American Educational Research Journal*, 48(3), 794-831.

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, *42*(2), 371-406

Howell, P. B., Cook, C., & Faulkner, S. A. (2013). Effective middle level teaching: Perceptions on the preparedness of newly hired teachers. *Middle Grades Research Journal*, *8*(3), 1.

Hussar, W. & Bailey, T.M. (2018). Projections of Education Statistics to 2026 (NCES 2018-019). U.S. Department of Education, Washington, DC: National Center for Education Statistics.

Ingersoll, R., Merrill, L. & May, H. (2014). What Are the Effects of Teacher Education and Preparation on Beginning Teacher Attrition?. *CPRE Research Reports.* Retrieved from https://repository.upenn.edu/cpre_researchreports/78

Ingersoll, R. & Perda, D. (2009). The Mathematics and Science Teacher Shortage: Fact and Myth. The Consortium for Policy Research in Education. Madison, WI.

Jackson, C.K. (2012). School competition and teacher labor markets: Evidence from charter school entry in North Carolina. *Journal of Public Economics*, 96(5-6), 431-448.

Jennings, J., & DiPrete, T. A. (2010). Teacher effects on social and behavioral skills in early elementary school. *Sociology of Education*, *83*(2), 135-159.

Johnson, S. (2018). These States are Leveraging Title II of ESSA to Modernize and Elevate the Teaching Profession. Center for American Progress. Retrieved from: https://cdn.americanprogress.org/content/uploads/2018/02/02103704/StatesLeveragingESSA-brief.pdf

Kane, T. J., Rockoff, J. E., & Staiger, D. O. (2008). What does certification tell us about teacher effectiveness? Evidence from New York City. *Economics of Education review*, *27*(6), 615-631.

Kane, T. J., Taylor, E. S., Tyler, J. H., & Wooten, A. L. (2011). Identifying effective classroom practices using student achievement data. *Journal of human Resources*, *46*(3), 587-613.

Keller, M. M., Neumann, K., & Fischer, H. E. (2017). The impact of physics teachers' pedagogical content knowledge and motivation on students' achievement and interest. *Journal of Research in Science Teaching*, *54*(5), 586-614.

Kelly, S. (2016). Shared Principles of Causal Inference. In D. Wyse, N. Selwyn, E. Smith, & Suter, L. E. (Eds.)., *The BERA/SAGE Handbook of Educational Research* (pp. 90-115). Sage.

Kelly, S. & Caughlan, S. (2011). The Hollywood teachers' perspective on authority, *Pedagogies: An International Journal*, 6:1, 46-65,

Kelly, S., & Northrop, L. (2015). Early career outcomes for the "best and the brightest" selectivity, satisfaction, and attrition in the beginning teacher longitudinal survey. *American Educational Research Journal*, *52*(4), 624-656.

Kennedy, M. M. (2008). Sorting out teacher quality. *Phi Delta Kappan*, *90*(1), 59-63.

Kennedy, M. (2016). Parsing the practice of teaching. *Journal of Teacher Education*, *67*(1), 6-17.

Kileen, K., Loeb, S., & Williams, I. (2015), A double draw of proximity: The importance of geography in teacher education and hiring decisions.

Kim, K. R., & Seo, E. H. (2018). The relationship between teacher efficacy and students' academic achievement: A meat-analysis. Social Behavior and Personality: an international journal 46(4), 529-540.

Kini, T., & Podolsky, A. (2016). Does teaching experience increase teacher effectiveness. *A Review of the Research*. Learning Policy Institute. Retrieved December 12, 2018 from: http://mrbartonmaths.com/resourcesnew/8.%20Research/Improving%20Teaching/Teaching%20Experience.pdf

Kleickmann, T., Richter, D., Kunter, M., Elsner, J., Besser, M., Krauss, S., & Baumert, J. (2013). Teacher's content knowledge and pedagogical content knowledge: The role of structural differences in teacher education. Journal of teacher education, 64(1), 90-106.

Koedel, C., & Betts, J. R. (2007). Re-examining the role of teacher quality in the *educational production function*. National Center on Performance Incentives, Vanderbilt, Peabody College.

Koedel, C., Mihaly, K., & Rockoff, J. E. (2015). Value-added modeling: A review. Economics of Education Review, 47, 180-195.

Koedel, C., Parsons, E., Podgursky, M., & Ehlert, M. (2012). *Teacher Preparation Programs and Teacher Quality: Are There Real Differences Across Programs?* CALDER Working Paper

79. Accessed August 29, 2015 from: http://www.caldercenter.org/sites/default/files/WP-79.pdf

König, J. (2013). First comes the theory, then the practice? On the acquisition of general pedagogical knowledge during initial teacher education. *International Journal of Science and Mathematics Education*, *11*(4), 999-1028.

Kukla-Acevedo, S. (2009). Leavers, movers, and stayers: The role of workplace conditions in teacher mobility decisions. *The Journal of Educational Research*, *102*(6), 443-452.

Kukla-Acevedo, S., Streams, M.& Toma, E. (2009), Evaluation of Teacher Preparation Programs: A Reality Show in Kentucky. University of Kentucky, Institute for Federalism and Intergovernmental Relations. No. 2009-09. Accessed July 20, 2015 from: http://www.ifigr.org/publication/ifir_working_papers/IFIR-WP-2009-09.pdf

Ladd, H. F. (2008, November). Value-added modeling of teacher credentials: Policy implications. In second annual CALDER research conference,“The Ins and Outs of Value-Added Measures in Education: What Research Says,” Washington, DC, November (Vol. 21).

Ladd, H. F., Clotfelter, C.T., & Holbein, J.B. (2015). The growing segmentation of the charter school sector in North Carolina. *Education Finance and Policy*, 12(4), 536-563.

Ladd, H. F., & Sorensen, L. C. (2015). Do Master's Degrees Matter? Advanced Degrees, Career Paths, and the Effectiveness of Teachers. Working Paper 136. National Center for Analysis of Longitudinal Data in Education Research (CALDER).

Ladd, H. F., & Sorensen, L. C. (2017). Returns to teacher experience: Student achievement and motivation in middle school. *Education Finance and Policy*, *12*(2), 241-279.

Lankford, H., Loeb, S., McEachin, A., Miller, L. C., & Wyckoff, J. (2014). Who enters teaching? Encouraging evidence that the status of teaching is improving. *Educational Researcher*, *43*(9), 444-453.

Ledwell, K., & Oyler, C. (2016). Unstandardized responses to a “standardized” test: The edTPA as gatekeeper and curriculum change agent. *Journal of Teacher Education*, *67*(2), 120-134.

Levine, A. (2006). Educating school teachers. *Education Schools Project*. Retrieved from: https://files.eric.ed.gov/fulltext/ED504144.pdf

Lewis, W. D., & Young, T. V. (2013). The politics of accountability: Teacher education policy. *Educational policy*, *27*(2), 190-216.

Luke, D. (2004). Multilevel Modeling. SAGE Publications.

Mason, K., Stanislawski, D., & Whitaker, D. (2019). Developing Embedded Signature Assessments for Continuous Improvement in Teacher Education. The Teacher Educator. 54(2), 171-192.

McDermott, P., & Gormley, K.A. (2016). Teachers' use of technology in elementary reading lessons. Reading Psychology, 37(1), 121-146.

Mertens, D. M. (2005). Research and evaluation in education and psychology: Integrating diversity with quantitative, qualitative, and mixed methods. 2nd Edition. Sage publications.

Metzler, J., & Woessmann, L. (2012). The impact of teacher subject knowledge on student achievement: Evidence from within-teacher within-student variation. *Journal of Development Economics, 99*(2), 486-496.

Mihaly, K., McCaffrey, D., Sass, T. R., & Lockwood, J. R. (2013). Where you come from or where you go? Distinguishing between school quality and the effectiveness of teacher preparation program graduates. *Education Finance and Policy*, *8*(4), 459-493.

Murnane, R. J., & Willett, J. B. (2011). Methods matter: Improving causal inference in educational and social science research. Oxford University Press.

Nardi, P.M. (2018). Doing survey research: A guide to quantitative methods. Routledge.

National Board for Professional Teaching Standards. (2019). In your state: North Carolina. Arlington, VA. Retrieved from: https://www.nbpts.org/in-your-state/in-your-state/nc/

National Board for Professional Teaching Standards. (2018). North Carolina State Profile. Retrieved from: http://www.nbpts.org/wp-content/uploads/State-Profiles_NC.pdf

Noell, G. H., Burns, J. M., & Gansle, K. A. (2018). Linking Student Achievement to Teacher Preparation: Emergent Challenges in Implementing Value Added Assessment. *Journal of Teacher Education*, 70 (2), 128-138.

North Carolina Department of Public Instruction (2016). North Carolina P-20W State Longitudinal Data System Grant: Project Narrative. Retrieved from: http://www.ncpublicschools.org/docs/data/ncp-20w/p20w-narrative.pdf

North Carolina Department of Public Instruction (2018a). Professional Educator's Licensure. Retrieved from: http://www.dpi.state.nc.us/licensure/

North Carolina Department of Public Instruction (2018b). 2016-2017 Report to the North Carolina General Assembly: Annual Report on the State of the Teaching Profession in North Carolina. Report 70.

North Carolina Department of Public Instruction (2018c). Final North Carolina ESSA State Plan. Retrieved from: https://files.nc.gov/dpi/documents/succeeds/nc-essa-state-plan-final.pdf

North Carolina Department of Public Instruction (2019). Percentage of Public School Students in Membership at Charter Schools. Retrieved from: https://files.nc.gov/dpi/documents/fbs/resources/data/csmembersmap.pdf

North Carolina Department of Public Instruction (2020a). Highlights from the North Carolina Public School Budget. Retrieved from: https://files.nc.gov/dpi/documents/fbs/resources/data/highlights/2020highlights.pdf

North Carolina Department of Public Instruction (2020b). Report to the North Carolina General Assembly: 2019 Annual Charter School Report Public School Budget. Retrieved from: https://files.nc.gov/dpi/documents/charterschools/resources/report-charter-schools-annual-report-2.15.2020.pdf

North Carolina Department of Public Instruction (2020c). Compensation for Public School Employees. Retrieved from: http://www.dpi.nc.gov

North Carolina Education Research Data Center (2019). *Data Housed at the North Carolina Research Data Center*. Retrieved from: https://childandfamilypolicy.duke.edu/wp-content/uploads/2013/08/DataHoused-at-NCERDC_Mar29.pdf

North Carolina General Assembly. (1985). North Carolina general statutes: Chapter 115C: Elementary and secondary education. Retrieved from: https://www.ncleg.net/EnactedLegislation/Statutes/HTML/ByArticle/Chapter_115C/Article_20.html

North Carolina General Assembly. (2017a). Excellent Educator in Every Classroom: Senate Bill 599. Retrieved from: https://www.ncleg.gov/Sessions/2017/Bills/Senate/PDF/S599v7.pdf

North Carolina General Assembly. (2017b). North Carolina Teaching Fellow: Senate Bill 252. Retrieved from: https://www.ncleg.gov/Sessions/2017/Bills/Senate/PDF/S252v0.pdf

Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects?. Educational evaluation and policy analysis, 26(3), 237-257.

Ozek, U., Carruthers, C., & Holden, K. (2018). Teacher Value-Added in Charter Schools and Traditional Public Schools. Working Paper 183. National Center for Analysis of Longitudinal Data in Education Research (CALDER).

Papay, J. P., Bacher-Hicks, A., Page, L.C., & Marinell, W.H. (2017). The challenge of teacher retention in urban schools: Evidence of variation from a cross-site analysis. Educational Researcher, 46(8), 434-448.

Papay, J. P., & Kraft, M. A. (2015). Productivity returns to experience in the teacher labor market: Methodological challenges and new evidence on long-term career improvement. *Journal of Public Economics*, *130*, 105-119.

Polly, D., (2014). Elementary school teachers' use of technology during mathematics teaching. Computers in the Schools, 31(4), 271-292.

Preston, C. (2014). Relationships between Teacher Preparation and beginning Teacher Quality in North Carolina Public Universities. Retrieved from *ProQuest* Dissertations Publishing.

Preston, C. (2017). University-based teacher preparation and middle grades teacher effectiveness. *Journal of Teacher Education*, *68*(1), 102-116.

Rabe-Hesketh, S. & Skrondal, A. (2012). Multilevel and longitudinal modeling using Stata. STATA Press.

Ramirez, G., Chang, H., Maloney, E.A., Levine, S.C., & Beilock, S.L. (2016). On the relationship between math anxiety and match achievement in early elementary school: The role of problem solving strategies. Journal of experimental child psychology. 141, 83-100.

Reardon, S. F. (2016). School district socioeconomic status, race, and academic achievement. Stanford Center for Educational Policy Analysis. Retrieved from:

Rezsonya, N. Yu, B., & Drew, J. (2019). Teacher Retention: Identifying Areas to Improve Teacher Recruitment and Retention in North Carolina Public Schools. SMU Data Science Review, 2(2), 10.

Rice, J. K. (2003). Teacher quality: Understanding the effectiveness of teacher attributes. Economic Policy Institute, 1660 L Street, NW, Suite 1200, Washington, DC.

Rice, J.K.(2010). The Impact of Teacher Experience: Examining the Evidence and Policy Implications. Brief No. 11. National center for analysis of longitudinal data in education research.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, *73*(2), 417-458.

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American economic review*, *94*(2), 247-252.

Rockoff, J. E., Jacob, B. A., Kane, T. J., & Staiger, D. O. (2011). Can you recognize an effective teacher when you recruit one?. *Education finance and Policy*, *6*(1), 43-74.

Rockoff, J. E., Staiger, D. O., Kane, T. J., & Taylor, E. S. (2012). Information and employee evaluation: Evidence from a randomized intervention in public schools. *American Economic Review*, *102*(7), 3184-3213.

Ronfeldt, M. (2012). Where should student teachers learn to teach? Effects of field placement school characteristics on teacher retention and effectiveness. *Educational Evaluation and Policy Analysis*, *34*(1), 3-26.

Ronfeldt, M. (2014). Does pre-service preparation matter? Examining an old question in new ways. Teachers College Record, 116 (10), n10.

Ronfeldt, M. (2015). Field placement schools and instructional effectiveness. *Journal of* Teacher Education, 66(4), 304-320.

Ronfeldt, M., Lankford, H., Loeb, S., & Wyckoff, J. (2011). How teacher turnover harms student achievement (NBER Working Paper No. 17176). Cambridge, MA: National Bureau of Economic Research

Ronfeldt, M., & Reininger, M. (2012). More or better student teaching?. *Teaching and* Teacher Education, 28(8), 1091-1106.

Ronfeldt, M., Reininger, M., & Kwok, A. (2013). Recruitment or preparation? Investigating the effects of teacher characteristics and student teaching. *Journal of Teacher Education*, *64*(4), 319-337.

Rutledge, S.A., Harris, D.N., Thompson, C.T.& Ingle, W.K. (2008). Certify, Blink, hire: An examination of the process and tools of teacher screening and selection. Leadership and Policy in Schools, 7(3), 237-263.

Sanders, W. L., Ashton, J. J., & Wright, S. P. (2005). Comparison of the Effects of NBPTS Certified Teachers with Other Teachers on the Rate of Student Academic Progress. Final Report. *National Board for Professional Teaching Standards*.

Santagata, R., Zannoni, C., & Stigler, J. W. (2007). The role of lesson analysis in pre-service teacher education: An empirical investigation of teacher learning from a virtual video-based field experience. *Journal of mathematics teacher education*, *10*(2), 123-140.

Sass, T. R. (2011). Certification Requirements and Teacher Quality: A Comparison of Alternative Routes to Teaching. Working Paper 64. National Center for Analysis of Longitudinal Data in Education Research.

Sawchuck, S. (2013a). Differing Viewpoints on Teacher Preparation Reflected in New Bills. Education Week. Retrieved from: https://blogs.edweek.org/edweek/teacherbeat/2013/05/Differing_viewpoints_on_teacher_prep.html

Sawchuck, S. (2013b) N.C. Teachers to Lose Tenure Salary Bumps. Education Week. 37(32), 4. Retrieved from: https://www.edweek.org/ew/articles/2013/08/07/37brief-5.h32.html

Sawchuck, S. (2014). Teacher Pay Raise a Mixed Deal in North Carolina. Education Week. Retrieved from: https://www.edweek.org/ew/articles/2014/08/13/01carolina.h34.html

Singer, J. D., & Willett, J. B. (1993). It's about time: Using discrete-time survival analysis to study duration and the timing of events. *Journal of educational statistics, 18*(2), 155-195.

Stieha, V., Shadle, S.E., & Paterson, S. (2017). Stirring the pot: Supporting and challenging general education science, technology, engineering, and mathematics faculty to change teaching and assessment practice. The Journal of General Education, 65(2), 85-109.

Tatto, M. T., Savage, C., Liao, W., Marshall, S. L., Goldblatt, P., & Contreras, L. M. (2016). The emergence of high-stakes accountability policies in teacher preparation: An examination of the US Department of Education's proposed regulations. *education policy analysis archives, 24*, 21.

Teach for America (2018). Teach for America: What We Do. Retrieved from: https://www.teachforamerica.org/what-we-do

Tennessee Higher Education Commission. (2011). 2011 Report Card on the Effectiveness of Teacher Training Programs. State Board of Education. Nashville, TN. Retrieved from: https://www.tn.gov/content/dam/tn/stateboardofeducation/documents/teacherprep/RC2011_ETTP.pdf

Tennessee State Board of Education (2018). 2018 Educator Preparation Report Card. Nashville, TN. Retrieved from: https://teacherprepreportcard.tn.gov/

The State of North Carolina. (2010). Race to the Top Proposal, Phase 2. Retrieved from http://www2.ed.gov/programs/racetothetop/phase2- applications/index.html

Turner, H. M., Goodman, D., Adachi, E., Brte, J., & Decker, L. (2012). Evaluation of Teach For America in Texas schools. San Antonio, TX: Edvance Research, Inc. Retrieved April, 5, 2018.

United States Department of Education (2003), Meeting the Highly Qualified Teachers Challenge: The Secretary's Second Annual Report on Teacher Quality, Washington, D.C. Retrieved from: https://title2.ed.gov/Public/TitleIIReport03.pdf

United States Department of Education (2010). Race to the Top Appendix B: Scoring Rubric. Retrieved from: https://www2.ed.gov/programs/racetothetop/scoringrubric.pdf

United States Department of Education. (2011). Our Future Our Teachers. Washington, DC. Retrieved from: http://www.ed.gov/sites/default/files/our-future-our-teachers.pdf

United States Department of Education (2015a). Condition of Education—Teacher Turnover: Stayers, Movers, and Leavers. Retrieved from: https://nces.ed.gov/programs/coe/indicator_slc.asp

United States Department of Education (2015b). Table 213.20 Staff Employed in public elementary and secondary school systems by type of assignment and state or jurisdiction: Fall 2015. https://nces.ed.gov/programs/digest/d17/tables/dt17_213.20.asp?current=yes

United States Department of Education (2016a). Table 209.24: Number and percentage distribution of teachers in public elementary and secondary schools, by whether they entered teaching through an alternative route to certification program and selected teacher and school characteristics. Retrieved from: https://nces.ed.gov/programs/digest/d17/tables/dt17_209.24.asp

U.S. Department of Education (2016b). Teach For America. Retrieved from: https://ies.ed.gov/ncee/wwc/Docs/InterventionReports/wwc_tfa_083116.pdf

U.S. Department of Education (2016c). 2016 Title II Reports, National Teacher Preparation *Data: North Carolina.* Retrieved from https://title2.ed.gov/Public/Home.aspx

United States Department of Education (2016d). Teacher Preparation Issues. Final Regulations. Federal Register,81, 210. Retrieved from: https://www.federalregister.gov/documents/2016/10/31/2016-24856/teacher-preparation-issues

United States Government Accountability Office (2015). Teacher Preparation Programs: Report to the Ranking Member, Subcommittee on Health, Employment, Labor, and Pensions, Committee on Education and the Workforce, House of Representatives (GAO-15-598). Retrieved from http://www.gao.gov/assets/680/671603.pdf

United States Department of Labor (2017). Occupational Employment Statistics: Data Tables for the overview of May 2017 occupational employment and wages. Washington DC. Retrieved from: https://www.bls.gov/oes/2017/may/featured_data.htm#largest4

United States Department of Labor (2019). Occupational Employment Statistics: Data Tables for the overview of May 2019 occupational employment and wages. Washington DC. Retrieved from: https://www.bls.gov/oes/2019/may/featured_data.htm#largest4

Van der Sandt, S., & O'Brien, S. (2017). Impact of Instructor Teaching Style and Content Course on Mathematics Anxiety of Preservice Teachers. Journal of Technology Education, 29(1), 95-111.

von Hippel, P. T., Bellows, L., Osborne, C., Lincove, J. A., & Mills, N. (2016). Teacher quality differences between teacher preparation programs: How big? How reliable? Which programs are different? *Economics of Education Review*, *53*, 31-45.

von Hippel, P. T., & Bellows, L. (2018). How much does teacher quality vary across teacher preparation programs? Reanalyses from six states. *Economics of Education Review*. 64, 298-312.

Voss, T., Kunter, M., & Baumert, J. (2011). Assessing teacher candidates' general pedagogical/psychological knowledge: Test construction and validation. *Journal of educational psychology*, *103*(4), 952.

Wayne, A. J., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. Review of Educational research, 73(1), 89-122.

Wiswall, M. (2013). The dynamics of teacher quality. *Journal of Public Economics*, 100, 61-78.

Woltman, H., Feldstain, A., MacKay, J. C., & Rocchi, M. (2012). An introduction to hierarchical linear modeling. Tutorials in Quantitative Methods for Psychology, 8(1), 52-69.

Xu, Z., Hannaway, J., & Taylor, C. (2011). Making a difference? The effects of Teach For America in high school. Journal of Policy Analysis and Management, 30(3), 447-469.

Youngs, P., & Qian, H. (2013). The influence of university courses and field experiences on Chinese elementary candidates' mathematical knowledge for teaching. *Journal of Teacher Education*, *64*(3), 244-261