# Early Career Teaching Supports, Instructional Growth, and Employment Decisions

by

Emanuele Bardelli

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Educational Studies)
in the University of Michigan
2022

Doctoral Committee:

Associate Professor Matthew Ronfeldt, Chair
Professor Deborah Ball
Professor Susan Dynarski, Harvard University
Professor Brian Jacob

Emanuele Bardelli
bardelli@umich.edu
ORCID iD: 0000-0003-3383-9315

# DEDICATION

*To Lauren*

## ACKNOWLEDGMENTS

This dissertation has benefited of feedback of multiple people, without whom this work would not have been possible.

First, I am deeply indebted to Matt Ronfeldt for the countless hours that we have spent together discussing this set of studies. His insights have made the papers in this dissertation much better. Also, I am grateful for his patience in helping with line editing my writing and in pushing me to always clarify my thinking.

I am also grateful for the feedback on prior drafts of this dissertation from the members of my committee. I am thankful for your grace in pointing out shortcomings in these prior drafts and in pushing me to make my work better. Deborah Ball has been instrumental in developing and refining the theoretical framing for this work. Sue Dynarski has provided feedback and encouragement on this project, on wonky ideas during its inception, on various iterations are it progressed, and finally on crafting policy and practice. Brian Jacob has lent a critical eye to the methods and results sections, pushing me to better explain and sharpening my reasoning.

I have learned all the research skills I used to complete this dissertation as part of my research apprenticeships. My work with Matt Ronfeldt's research group has been pivotal in my development as a researcher, from teaching me the basics data management skills, to designing and carrying out complex research designs, to reporting results in conference presentations and journal publications. His research-practice partnership with the Tennessee Department of Education is also foundational to this work and has been a major inspiration for this dissertation. Two other research laboratories have welcomed and trained me: Pat Herbst's Grasping the Rationality of Instructional Practice (GRIP) lab and Matt Diemer's Advancing Critical Consciousness, Methods, & Equity (AC$^2$ME) lab. I am grateful of the opportunity to learn from both of these labs and to participate in their research activities.

During my research apprenticeships, I also had the pleasure to getting to know and work with outstanding graduate students, many of whom became friends and collaborators. I am grateful for Matt Truwit for be a critical collaborator, slack buddy, and for providing feedback on early drafts of this dissertation. Inah Ko, Nic Boileau, and Mollee Shultz were instrumental during my first years of graduate school and gave me the psychometrics bug. Amanda Weissman, Mike Ion, and

Michael Frisby have shared many coffee breaks with me. I appreciated our conversations about everything and anything.

There are numerous other teachers, faculty members, and students who have been influential over the years. Though I am unable to mention everyone by name, I am thankful for your support and encouragement. I would not been here without all of your help. I will try to pay it forward to others when I have the chance.

I am also thankful to the Tennessee Department of Education and the Tennessee Education Research Alliance for providing access to the data and computational resources for this work. I would have not be able to conduct these studies without access to the wealth of educational data from Tennessee.

This dissertation would not have been possible without the generous financial support from the Institute of Education Sciences (IES), the U.S. Department of Education through a pre-doctoral training grant (PR/Award R305B150012), and the University of Michigan's School of Education one-term dissertation finishing grant.

# TABLE OF CONTENTS

# LIST OF FIGURES

FIGURE

# LIST OF TABLES

TABLE

# ABSTRACT

This three-paper dissertation studies the relationship between teaching supports and early-career teachers' instructional growth and employment decisions, including to migrate schools or leave teaching. Using extensive survey and administrative data, I develop latent measures for teaching supports that I use to explore their relationships with measures of early-career teachers' instructional effectiveness and employment decisions.

In Paper 1, I analyze the associations among the instructional support measures and instructional effectiveness. I find that reporting higher levels of teaching supports are associated with faster improvement in observation ratings but not teacher value-added measures. Professional learning & development supports and evaluative feedback have stronger associations with growth than the other types of supports. Working conditions have the lowest among them. Early career teachers who report teaching supports one standard deviation above the average become eligible for tenure one year sooner than their peers (three vs. four years on average). These results suggest that high levels of teaching supports, particularly ones that are sustained over many years and targeting early career teachers, have promise for supporting instructional improvement.

In Paper 2, I study which teaching supports for early career teachers relate to early career teachers' employment decisions, including moving schools, moving districts, or leaving teaching. Furthermore, I study the extent to which the relationships between teaching supports and employment decisions is influenced by changes in instructional effectiveness. I find that, while teaching supports are associated with reductions in all forms of turnover, the magnitudes are greatest regarding leaving the profession and moving districts as compared with migrating schools within the same district. Improvement in instructional effectiveness due to teaching supports mediates about twenty percent of this relationship. These results suggest that investment in teaching supports improves teacher retention, both though direct effects on teachers' employment decisions and instructional improvement that results in greater teacher retention.

In Paper 3, I describe the development of the teaching supports measures and assess their variation across different kinds of teachers, schools, and districts. I find that my measures of teaching supports have adequate psychometric properties. More within-school rather than between-school variation in terms of average teaching supports offered to teachers, as I observe the majority of variation in teaching supports between-teachers within the same school. Finally, some variation in

these scores along teacher and school characteristics. Teachers in their first three years of teaching report higher levels of teaching supports than more experienced teachers while STEM teachers and teachers working in schools that have larger enrollment or that serve more students eligible for free or reduced price lunch report lower level of supports than other teachers.

This dissertation informs both policy decisions around beginning teachers and opens new directions for future research. My results suggest that early career teachers should receive high levels of teaching supports sustained for at least three years in order to be able to observe substantive returns. Programs aimed to support early career teachers might have the best chance for success when they are implemented by school districts or state-wide rather than at the school level. Given that teaching supports are more strongly associated with reduction in leaving the profession and migrating districts than migrating schools within districts, state and district leaders are more likely than school leaders to benefit from investment in teaching supports.

# CHAPTER 1

# Dissertation Overview

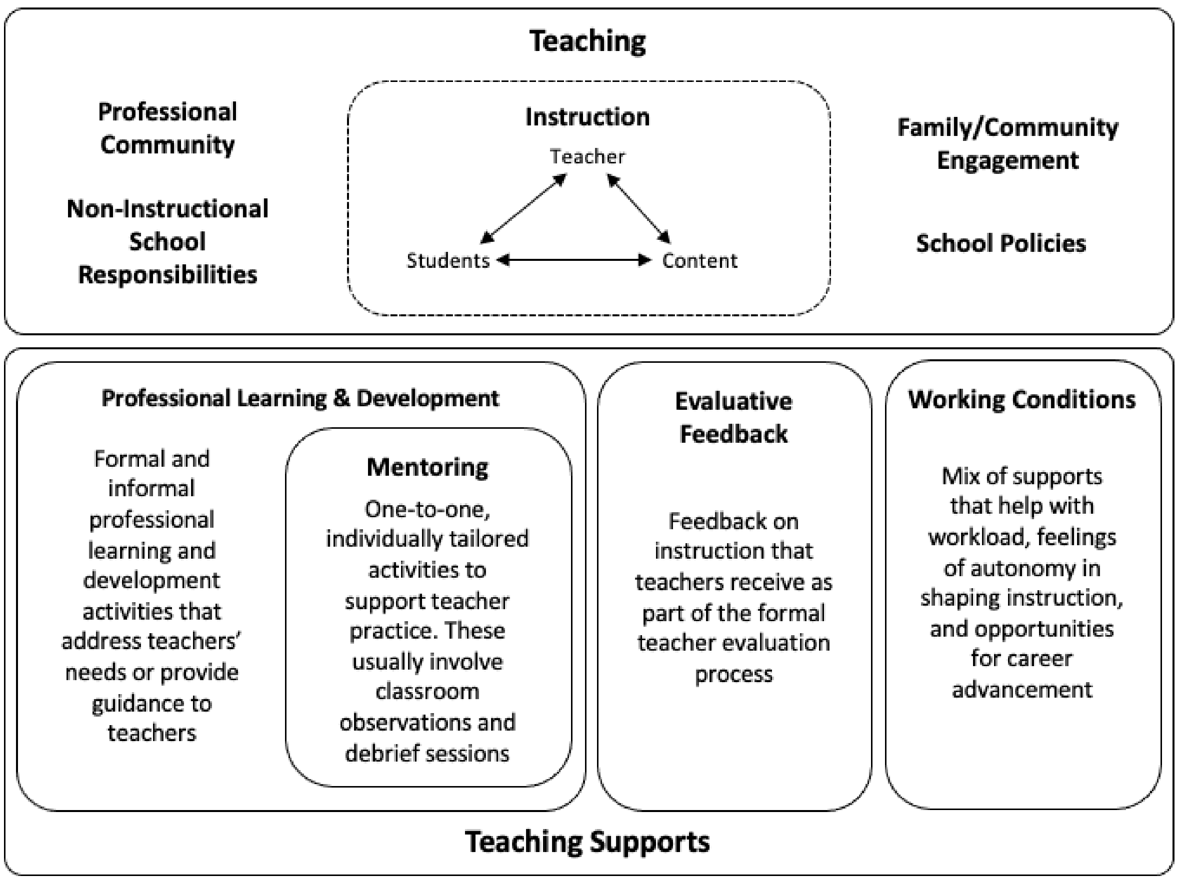
The first years of teaching are the defining period for any teacher's career. During this time, teachers transition from their own schooling experiences and teacher preparation programs to their new roles as teachers. Multiple scholars have reported on the challenges new teachers face in this transition and on ways to ameliorate these challenges. Yet, at least one-third of new teachers leave the profession within their first five years (Ingersoll & Merrill, 2012). For this reason, teacher induction programs, professional development programs, or other ways to support early career teachers have become more and more popular over the past decades (Howe, 2006; Ingersoll & Smith, 2004; Ingersoll & Strong, 2011) and are discussed as a possible policy lever to help address the teacher shortage following the COVID-19 pandemic. However, the evidence of the benefits of these support programs is mixed at best, with some work finding positive relationship between participation in teacher induction and teacher employment decisions (Ingersoll et al., 2014; Kang & Berliner, 2012; Ronfeldt & McQueen, 2017), while others have found null effects (Glazerman et al., 2010; Isenberg et al., 2009). This dissertation project adds to this growing body of work and focuses on how types of teaching supports relate to early career teachers' instructional effectiveness and employment decisions, including mobility between schools and teacher attrition. In addition, I describe the technical details of the development and validation of the teaching supports measures. As a result, this work promises to inform policy on the best ways to support the professional growth and retention of early career teachers.

Building on previous work suggesting that teaching supports for new teachers play a crucial role in the development of new teachers' practice and in their employment decisions (Boyd et al., 2008; Bruno et al., 2019; Papay & Kraft, 2015), I designed and conducted a set of studies with the goal of describing the types of teaching supports available to early career teachers and their connection to instructional improvement, student outcomes, and teacher employment decisions. The work in organizational science, school improvement, and teacher induction provides a road-map of the kinds of teaching supports that have been shown to improve early career teachers' classroom practice, student achievement, and teacher retention (Bryk et al., 2009; Ingersoll & Strong, 2011; Moore Johnson, 2020; Ronfeldt & McQueen, 2017).

## Teaching and Teaching Supports

This dissertation project ties together teaching and teaching supports, with the goal to observe how teaching support experiences influence teachers' instructional effectiveness outcomes and employment decisions. It is necessary, therefore, to clarify what I mean by teaching and how teaching is connected to its supports. Figure 1.1 summarizes visually how I conceptualize the relationship between teaching and teaching supports.

Figure 1.1: Teaching and Teaching Supports




*Teaching*, as used in everyday language, has two complementary meanings (Herbst & Chazan, 2016). On one hand, teaching is used to describe activities that produce learning. In this definition, a person's actions are closely tied with learning outcomes, but it is not necessarily tied to teaching happening in schools (e.g., teaching someone how to fish). On the other hand, teaching can be defined simply as what teachers do. In this case, teaching is separated from its outcomes and is centered on the day-to-day practices and operations that separate teaching from other professions. In short, teaching becomes everything that teachers do regardless of its connection to learning or students. As Cohen (2011) pointed out, these two competing definitions of teaching are problematic in their own terms. As a result, he proposes a third definition of teaching that combines these two everyday uses of teaching in a single definition more suited to education research. In this case, teaching becomes *the work that teachers do that is deliberately oriented to produce learning*. I adopt this definition as the foundation of this dissertation. Moreover, I follow Herbst and Chazan (2016) in sharpening Cohen's original definition with the goal to separate deliberate teaching practices that could directly produce learning from other practices and actions that could only indirectly influence learning. The former is what is often referred to as *instruction* or the relationship between teachers, students, and content in environments (Cohen et al., 2003). The latter becomes teacher actions and practices that happen in the environments surrounding instruction, such as their engagement with their schools' professional community, engaging with families or the community, implementing school policies, or other non-instructional school responsibilities outside the classroom.

Multiple scholars have shown that teaching can be supported in various ways (Bryk et al., 2009; Moore Johnson, 2020). In this dissertation I foreground three teaching supports that have the potential to impact teaching and result in improved student outcomes: *Professional Learning & Development*, *Evaluative Feedback*, and *Working Conditions*.

Professional learning & development activities includes both formal and informal learning opportunities for teachers that are meant to address their professional needs or to provide them with guidance. These activities can influence instruction when their focus is directly tied with parts of the instructional triangle. In fact, Lynch et al. (2019) conducted a meta-analysis on the effects of professional development on STEM instruction and found that professional development that focused on curricular material use (i.e., teacher and content), content-specific formative assessment (i.e., students and content), and pedagogical content knowledge or how students learn (i.e., teacher and students) had the highest moderating effects on professional development effectiveness. At the same time, it is possible for professional development to impact other aspects of teaching that are not directly tied to instruction. For example, a workshop focused on how to collaborate with colleagues in small learning communities might not be directly tied to instruction but could have an indirect effect on instruction by changing the environment around instruction.

Within professional learning & development, I purposefully isolate mentoring activities. Separating mentoring from other forms of professional development is important because mentoring is one major components of most new teacher induction activities (Kang & Berliner, 2012; Ronfeldt & McQueen, 2017; Smith & Ingersoll, 2004). In this case, mentoring includes one-on-one interactions between a teacher and a colleague with the specific goal to address the teacher's needs.

Similar to general professional development activities, mentoring can have a direct impact on instruction when focused on specific parts of the instructional triangle (e.g., mentoring on how to best launch a lesson) or it can more generally impact environmental features that surround instruction (e.g., how to deal with a difficult parent).

Another aspect of learning to teach is to become familiar with the expectations of teaching. This part of teaching has become more salient over the past two decades with the introduction of formal teacher evaluation systems that are often used for high-stakes employment decision, including in the state that is the focus of this dissertation. For this reason, I focus on the feedback that teachers receive as part of their formal teacher evaluation process. Moreover, both policy documents and researchers have called for using teaching evaluation experiences as an opportunity for feedback on teaching practice with the goal to improve instruction (Papay, 2012; U.S. Department of Education, 2009).

Finally, one of the most salient features of the environments where instruction happens is working conditions. Working conditions envelop different aspects of the school environment that make instruction less (or more) burdensome by taking up teachers' time, improving teachers' feelings of autonomy, or by opening up paths for career improvement. Working conditions, while most likely not directly related to instruction, can be important in shaping an environment where teachers can improve and thrive (Bryk et al., 2009; Moore Johnson, 2020).

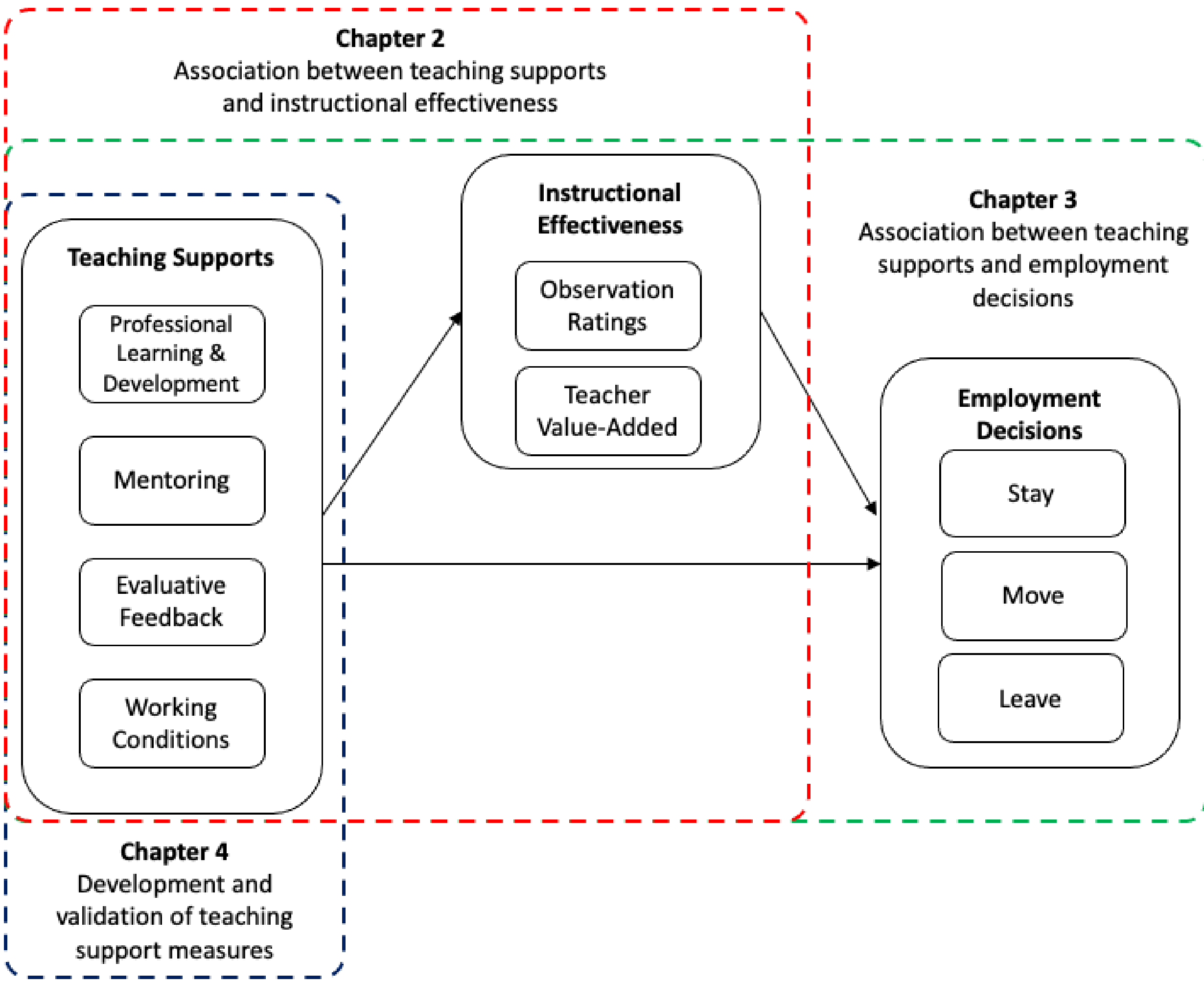
## Dissertation Overview

Figure 1.2 shows how the three chapters in this dissertation are related to each other. Scanning the figure from left to right, I start with the types of teaching supports that early career teachers are likely to receive. Using extensive survey and administrative data, I develop latent measures for *professional learning & development*, *mentoring*, *evaluative feedback*, and *working conditions*. I then use these latent measures to explore their relationships with measures of early career teachers' instructional effectiveness and employment decisions. In the sections below, I will describe each chapter in more detail.

In Chapter 2, I analyze the associations among the instructional support measures and measures of early career teachers' instructional effectiveness (i.e., observation scores and value-added to student test scores measures). The goal of this chapter is to understand which of these teaching supports are related to early career teaching performance (both initial levels and subsequent growth). Additionally, I test various model specifications using different sources of variation in the measures of teaching supports to estimate plausibly causal effects of these teaching supports measures on the growth in observation scores and teacher value added estimates.

In Chapter 3, I study which teaching supports for early career teachers relate to early career

Figure 1.2: Relationship Between the Papers in This Dissertation



teachers' employment decisions, including moving schools, moving districts, or leaving teaching. Furthermore, I study the extent to which the relationship between teaching supports and employment decisions is influenced by teachers' instructional effectiveness.

In Chapter 4, I describe the technical details behind the development of the teaching supports measures and assess their variation across different kinds of teachers, schools, and districts. These teaching supports include supports for *professional learning & development*, *mentoring*, *evaluative feedback*, and *working conditions*. As discussed above, these teaching supports measures capture different dimensions of the environment in which early career teachers work. For each measure, I use indicators from teacher surveys to develop and validate latent measures for each dimension. The measure development follows a two-step approach based on psychometric best practices, including exploratory analyses to identify whether and how survey items relate to each instructional support type and confirmatory analyses to ensure the validity of these measures. In the second part of this paper, I explore variation in these measures for teacher, school, and district characteristics.

# Contributions

This dissertation makes several contributions to our knowledge of how teaching supports relate to instructional effectiveness and employment decisions. This dissertation is one of the first to link state-wide survey data on teaching supports to workforce outcomes. Prior research has either relied on national samples of teachers (e.g., Kang and Berliner, 2012; Ronfeldt and McQueen, 2017; Smith and Ingersoll, 2004) or a selected group of teachers from a few school districts (e.g., Auletto, 2021; Redding et al., 2019; Smith et al., 2018). While the nationwide studies are nationally representative, they depend upon surveys with limited information about the kinds of supports teachers receive. In addition, these studies did not link teaching supports to measures of instructional effectiveness. On the contrary, studies with a selected group of teachers have better, more finely grained measures of teaching supports but lack the potential for generalization. My work builds on the strengths of this prior research and tries to address its limitations by developing comprehensive measures of teaching supports for a large, statewide sample of teachers. This process allows me to observe the full variation in early career teaching supports that are available to beginning teachers, as well as access to a rich administrative data that includes instructional effectiveness outcomes and employment decisions, which gives me the opportunity to develop nuanced measures of teaching supports.

This dissertation is also the first study to explore whether and to what extent instructional effectiveness measures mediate the relationship between teaching supports and employment decisions. To my knowledge, previous work has only studied the relationship between teaching supports and either instructional effectiveness or employment decisions separately. On one hand, being able to parse apart whether and how teaching supports predict teacher retention above and beyond their association with instructional effectiveness can help identify those supports that policymakers and practitioners alike can likely use to improve the retention of early career teachers. On the other hand, if the relationship between teaching supports and employment decisions is fully mediated by changes in instructional effectiveness, my results would suggest that focusing on teaching supports that develop early career teachers' instructional effectiveness will have the added benefit of also helping with teacher retention.

Finally, Chapter 4 reports the technical details behind the development and validation of my measures of teaching supports. As I rely on secondary data analysis of existing survey data, it is essential to assess the alignment between the survey items and the measures that I develop. Psychometrics offer a robust set of methods to develop and assess the alignment of survey items to latent constructs. Moreover, this is an improvement over previous work that has either counted the total number of teaching supports to develop measures of support intensity (Kang & Berliner, 2012; Ronfeldt & McQueen, 2017; Smith & Ingersoll, 2004) or relied on principal components factor analysis to reduce the dimensionality of a complex survey instrument (Auletto, 2021).

## Research Questions

The overarching research question for this project is: What kinds of teaching supports are available to early career teachers, and how do these supports relate to measures of their instructional effectiveness (i.e., observation ratings and value-added measures) and employment decisions (i.e., mobility and retention)? In the following three chapters, I consider the following set of more specific research questions:

CH2.RQ1. What is the association between overall teaching supports and instructional effectiveness outcomes, as measured by observational ratings and student achievement gains?

CH2.RQ2. Which types of teaching supports are related to instructional effectiveness outcomes?

CH3.RQ1. What is the relationship between different types of teaching supports and early career teachers' employment decisions?

CH3.RQ2. Does instructional effectiveness moderate the relationship between teaching supports and employment decisions?

CH4.RQ1. Which measures for types of teaching supports can be developed using state-wide survey data?

CH4.RQ2. What are the psychometric properties of these measures for types of teaching supports?

CH4.RQ3. How do these measures vary across teacher, school, and district characteristics?

## Findings Overview

Results from Chapter 2 suggest that teaching supports matter for early career teachers. I find that the observation ratings for teachers who report receiving higher than average (i.e., one standard deviation above the average) teaching supports grow (in terms of instructional effectiveness) at faster rates than their colleagues who do not. If this level of support is sustained over the early career period, it can result in teachers reaching the minimum requirement for tenure—an observation ratings of *above expectations*–in three years, which is a year earlier than their peers that only reported average levels of teaching supports. Additionally, teachers in their first three years of teaching experience appear to especially benefit from teaching supports, as the relationship between reporting receiving one standard deviation above the state average is almost a quarter larger for early career teachers than the other teachers. However, results for value-added to student test measures (VAMs) suggest that, while early career teachers' VAMs grow at faster rates when they experience increased teaching supports, results are less consistent across model specifications, suggesting that VAMs are likely less responsive to teaching supports than observation ratings.

Looking at employment decisions, Chapter 3 shows that teachers who receive more teaching supports are less likely to leave teaching or move to different schools. Teachers in their first three years of teaching appear to be especially responsive to increased teaching supports, with their retention rate improving by almost 2 percentage points or about 25 percent and their mobility rate decreasing by 1.7 percentage points or about 15 percent. Moreover, only about 20 percent of this relationship is mediated by changes in instructional effectiveness, suggesting that teacher employment decisions are less tied to how effective a teacher is and more about other aspects of teaching supports. In other words, employment decisions are related to more than just how instructionally effective a teacher is and are responsive to teaching supports above and beyond their relationship with instructional improvement.

Finally, Chapter 4 reports on the measure development and validation processes and shows that it is possible to use large state-wide survey data to develop teaching support measures using a longitudinal survey of teachers not specifically designed to collect data on these constructs. All four of these measures demonstrate adequate psychometric properties to warrant their use in further analysis. Furthermore, results from heterogeneity in these measures show that about half of the variation in these measures is not explained by teacher, school, or district level variables, suggesting that there is enough year-to-year variation in these measures to use them in a longitudinal panel.

# CHAPTER 2

# Role of Teaching Supports in Early Career Teachers' Professional Growth

An analysis of recent trends in teaching workforce composition has found that teachers are becoming both "grayer" and "greener" (Ingersoll & Merrill, 2012). The first finding suggests that there might be an increase in open teaching positions as older, more experienced teachers retire (Sutcher et al., 2019). The second finding suggests that schools might need to adjust the kinds of professional development and other activities that they make available to their staff to target the needs of new, less experienced teachers. If this is the case, then policymakers and school administrators need to know which kind of supports are better suited to accelerate the development of new teachers. Traditionally, teacher induction programs (Howe, 2006; Ingersoll & Smith, 2004; Ingersoll & Strong, 2011) are one of the preferred ways to support new teachers during their first few years in the profession. These induction programs vary in both the kinds and quality of activities offered to teachers, but they all share the same design principle of providing targeted professional development for early career teachers. Observational studies using large, nationally representative samples of teachers found participating in induction activities benefited teachers, at least in terms of teacher retention (Henke et al., 2000; Kang & Berliner, 2012; Ronfeldt & McQueen, 2017; Smith & Ingersoll, 2004), and found mixed evidence for instructional improvement (Davis & Higdon, 2008; Fletcher et al., 2008; Glazerman et al., 2010; Isenberg et al., 2009; Kang & Berliner, 2012; Nevins Stanulis & Floden, 2009; Rockoff, 2008; Thompson et al., 2005). More recent, targeted studies (Redding et al., 2019; Smith et al., 2018) have explored how induction activities could be associated with improvements in teacher instructional outcomes for early career teachers between their first and third year of teaching. These studies have found that improvement in their instructional quality, as measured by trained observers, was not associated with the kinds of induction supports they received, even if early career teachers felt that induction activities helped them improve their teaching practice.

In this chapter, I use state-wide data from Tennessee to build upon prior work by studying how teacher instructional outcomes—as measured by instructional effectiveness outcomes including observation ratings and value-added to student test scores—are associated with teaching supports that early career teachers experience. This chapter makes two contributions to previous work on teacher induction. First, it is one of the first state-wide analyses, to my knowledge, linking teaching supports and instructional effectiveness. Prior work relied on nationally representative surveys, such as the School and Staffing Survey, or ad hoc samples of early-career teachers. State-wide data addresses limitations in these prior studies by focusing on a self-contained universe of teachers within a state.

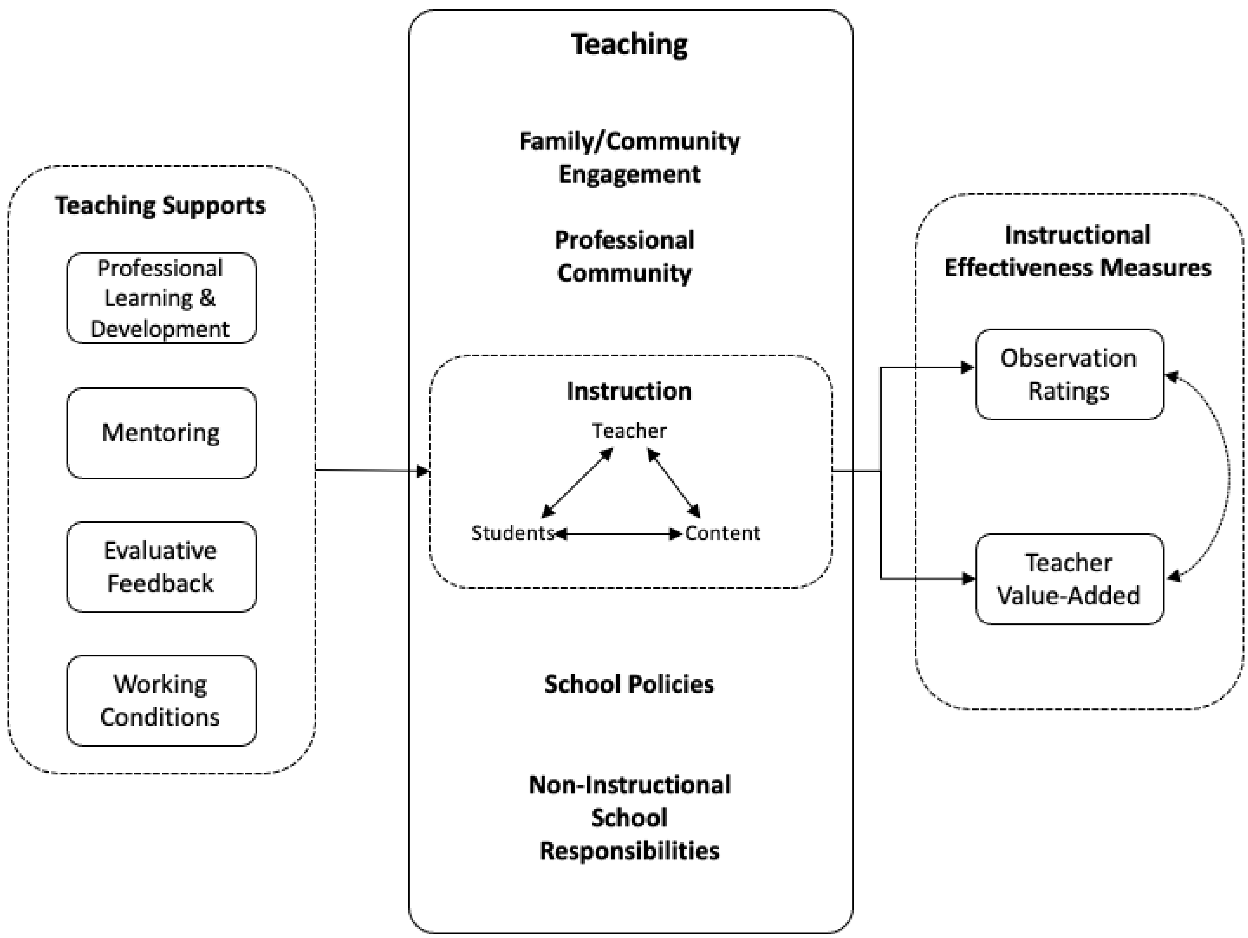
Second, this chapter observes teaching supports for early career teachers and the variation in these supports within and across teachers. I take advantage of a unique longitudinal survey program in Tennessee to develop measures of teaching supports that naturally occur in schools; by contrast, many prior studies are evaluations of specific teacher induction programs or interventions that bundle together in a black box the various formal and informal types of supports experienced across teachers and settings. These new measures of teaching supports allow me to focus on the experiences that early career teachers report and to observe their relationship with their instructional effectiveness outcomes. By leveraging both within and across teacher variation in teaching supports, I isolate the relationship between reported levels of teaching supports and changes in instructional effectiveness, an improvement on prior work that has relied on associations for cross-sectional samples of teachers.

## A Conceptualization of Teaching and Teaching Supports

This chapter centers teaching as the conduit between teaching supports and instructional outcomes (see Figure 2.1 for a visual representation). Following Cohen (2011) and Herbst and Chazan (2016), I define teaching as *the work that teachers do that is deliberately oriented to produce learning* and I separate instruction—as defined as the relationship between teacher, students, and content within environments (i.e., the instructional triangle, Cohen et al., 2003)—from other non-instructional activities that teachers might engage in as part of their daily work. This distinction is important because the outcomes of interest in this chapter are measures of instructional effectiveness, including observation ratings and teacher value-added measures, that are developed to provide a measure of instruction. For this reason, it is reasonable to hypothesize that teaching supports that directly target instruction (e.g., mentoring) might be more likely to result in changes in instructional effectiveness measures, while teaching supports that target other non-instructional activities (e.g., working conditions) might have a smaller relationship with changes in instructional effectiveness measures.

Teaching supports could lead to changes in instructional effectiveness measures either by directly affecting instruction—more likely, a part of the instructional triangle—or indirectly by shap-

Figure 2.1: Relationship Between Teaching Supports and Instructional Effectiveness Outcomes



ing the environments where instruction happens. In the first case, teaching supports that, for example, develop teachers' understanding of their curriculum can directly influence instructional change, leading to improvement in student outcomes. In fact, Lynch et al. (2019)'s meta-analysis found that professional development is more effective when its content directly relates to parts of the instructional triangle. In the second case, teaching supports can improve the environments where instruction happens, for example by facilitating teacher collaboration or by reducing teachers' non-instructional responsibilities, leading to a possible indirect improvement of instructional effectiveness measures.

Finding no relationship between teaching supports and instructional effectiveness would suggest, perhaps, that teaching practice naturally develops during the early career by new teachers exploring and discovering what works best for themselves, regardless of the teaching supports they experienced. This intuition could be in line with prior work that found no relationship between teacher induction practices and instructional effectiveness (Glazerman et al., 2010; Isenberg et al., 2009; Smith et al., 2018).

To further explore the relationship between teaching supports and instructional effectiveness, I start by identifying areas for teaching supports that have been linked to career development in prior literature. Four categories of teaching supports stand at the core of this conceptualization. These categories measure different types of teaching supports that early career teachers could experience during their first years in the teaching profession. These teaching supports include formal and informal activities that facilitate *professional learning & development*, *mentoring* supports with the goal to improve early career teachers' practice through classroom observations and feedback cycles, *evaluative feedback* supports that help early career teachers through the formal teacher evaluation process and make teaching expectations clear to them, and *working conditions* supports including collaboration with colleagues, opportunities for career advancement, and effective school leadership.

In the next sections, I review prior literature to elaborate on how each of these components is defined and conceptualized and summarize what we know about how each support is related to instructional effectiveness. It is important to note that categorizing teaching supports into separate components in this conceptualization is more of a helpful artifice than a systematic taxonomy of teaching supports. Though I define each support type as distinct from others, in practice these teaching supports interact and overlap with one another. This observation is most salient for the relationship between professional learning & development and mentoring, for which mentoring is often considered to be a sub-category of professional learning & development activities.

## Professional Learning & Development

Much has been written about the relationship between professional development and instructional outcomes (see, Borko, 2004; Desimone, 2009; Yoon et al., 2007 for reviews of this literature). In general, these reviews have found positive relationships between participation in professional development activities and improvement in student achievement and instructional effectiveness outcomes. Historically, researchers defined professional development as "any activity that is intended partly or primarily to prepare paid staff members for improved performance in present or future roles in the school districts" (Little, 1987, p. 491). This definition has evolved over time to also include informal opportunities beyond formal workshops or seminars from which teachers can learn and improve their practice. For example, Borko (2004) suggested that "for teachers, learning occurs in many different aspects of practice, including their classrooms, their school communities, and professional development courses or workshops. It can occur in a brief hallway conversation with a colleague, or after school when counseling a troubled child." (p. 4). This observation suggests the need to develop well-rounded measures of professional learning & development to account for the full extent of the learning opportunities that early career teachers might experience. Finally, studies that analyzed teacher induction practices using the School and Staffing Survey (e.g., Kang and Berliner, 2012; Ronfeldt and McQueen, 2017; Smith and Ingersoll, 2004) found that one of the most common induction supports for early career teachers is participating in professional development seminars, highlighting the importance that schools place on professional learning & development of early career teachers.

When focusing on the relationship between professional learning opportunities and early career teachers' development, the empirical evidence is mixed. While correlational studies found possible positive effects on student achievement for teachers participating in induction programs, the only randomized control trial on the effects of a set of comprehensive induction programs found consistent null effects across a variety of outcomes. For example, Thompson et al. (2005) studied the relationship between introducing the Beginning Teaching Support and Assessment (BTSA) program in California and changes in student achievement for teachers participating in the program. They reported positive associations across all six tested subjects between the intensity of the induction supports that teachers reported and student achievement scores. By contrast, the only randomized control trial assessing the effects of a beginner's seminars as part of a teacher induction program (Glazerman et al., 2010; Isenberg et al., 2009), found no effect between these professional development activities centered on various aspects of teaching practice and instructional outcomes, including classroom practices and student outcomes. Moreover, this program included up to two years of professional development activities for teachers in the treatment condition. At the same time, non-experimental results from the same study showed that teachers in the treatment had better student outcomes three years after the intervention (Glazerman et al., 2010), suggesting that the effects of professional development might take time to manifest in student outcomes.

Research on the effects of in-service mathematics teacher professional development suggested that teachers' mathematical knowledge for teaching (MKT, Ball et al., 2008) increases following professional development activities. In their meta-analysis of the effects of 326 professional development programs, Phelps et al. (2016) reported that mathematics knowledge for teaching scores improve between 0.16 and 0.26 standard deviation, on average, following professional development programs. At the same time, it appears that teachers' MKT scores do not develop during the first years of teaching in the absence of professional development activities. Desimone et al. (2016) followed 45 new teachers and measured their MKT scores during their first three years of teaching. They found that, while they improved their instructional quality scores, their MKT levels remained relatively stable during the same time. These results underscore the role of professional development as a teaching support in the growth of early career teachers' professional capacity.

## Mentoring

Beyond formal professional development activities, early career teachers usually receive mentoring about their teaching practice as part of their early career teaching supports (Desimone et al., 2013). These activities usually involve either a pairing of early career teachers with a more experienced peer, an instructional coach, or a supervisor with the goal of providing mentoring and advice to the beginning teacher. In a review of the literature on teacher induction, Luft and Cox (2001) found that most early career teachers received some form of mentoring, either through a formal induction program or informally at their school site. Surveys of induction programs have also shown that early career teachers are likely to receive some sort of instructional coaching or mentoring during their early careers (Ingersoll & Strong, 2011; Kang & Berliner, 2012; Ronfeldt & McQueen, 2017). It is less clear, however, how mentoring affects the development of instructional effectiveness for early career teachers. Alignment of mentoring with development goals seems to be an essential component in supporting the development of early career teachers' practice (Hammerness & Matsko, 2012).

Mentoring can develop both early career teachers' instructional practices and improve their student test score outcomes. Multiple studies have shown that receiving mentoring positively predicts instructional effectiveness (Davis & Higdon, 2008; Evertson & Smithey, 2000; Nevins Stanulis & Floden, 2009) and on student achievement (Fletcher & Strong, 2009; Fletcher et al., 2008; Rockoff, 2008; Thompson et al., 2005). For example, Fletcher et al. (2008) studied the association between variation in mentoring and student achievement from three districts in California. They found that teachers in districts with more intensive mentoring models had higher student test score gains than their colleagues in districts with less intensive mentoring. In a follow-up analysis, the authors followed the same group of early career teachers for five years and found no differences in student achievement gains between early career teachers who received mentoring and experienced teachers at the same schools, suggesting that mentoring supports could help with faster development of teaching practice among early career teachers.

Rockoff (2008) studied early career teachers' skill development as a result of participating in a mentoring program in New York City schools. Two main results emerge from this study. First, the authors found no differences in student achievement gains between new teachers without prior teaching experience who received mentoring and new-to-the-district teachers with prior teacher experience that did not receive mentoring. In addition, they reported a positive association between total hours of mentoring received and student achievement gains in both mathematics and reading. Together, these two results suggest that mentoring could help with faster instructional practice development, as it appears that inexperienced teachers who report receiving mentoring supports perform similarly to more experienced teachers, and that the intensity of mentoring supports appears to matter in this process.

This observation is also confirmed in other studies that focused on instructional coaching. Instructional coaching can be seen as a formalized version of mentoring that typically involves cycles of classroom observation, debriefing, and planning. Instructional coaching is often considered the best way to support the development of early career teachers, as these activities have been consistently shown improvement in both teachers' instructional practices and student achievement (Kraft et al., 2018). Recent work also shows that the format through which coaching is delivered, including online coaching (Kraft & Hill, 2020) or by peers (Burgess et al., 2019; Papay et al., 2020), leads to instructional improvement, suggesting that coaching activities—classroom observations paired with mentoring—support the development of early career teachers' professional capacity regardless of whether they are provided in person, online, or by a more expert peer.

Only recently have studies demonstrated causal evidence of the positive impacts of instructional coaching on teachers' practices and student achievement. Kraft and Hill (2020) reported on a randomized control trial for web-based coaching of middle school mathematics teachers. They found that participants who received coaching aligned with Common Core mathematics instruction improved both their instructional practices, as measured through the Mathematical Quality of Instruction (MQI) instrument and student surveys. On the other hand, they found no differences in the value-added to student achievement scores between the treatment and control groups. In another field experiment, Papay et al. (2020) matched pairs of high- and low-skilled colleagues and had them participate in a peer coaching program comprising peer observations and debrief sessions. They found that student achievement scores improved 0.12 standard deviation units for the low-skilled teacher in the pair.

## Evaluative Feedback

Both national and state policies have supported the development of teacher evaluation systems over the past twenty years. In 2002, No Child Left Behind act has mandated the development of a standardized assessment program across the United States. Performance on student assessments has been used to evaluate the effectiveness of individual teachers, schools, and districts. With the passage of Every Child Succeeds act in 2012, the federal government has supported the development and implementation of teacher evaluation systems based upon administrator classroom observation. These new teacher assessment instruments were meant to ameliorate the limitations of evaluating teachers using only student multiple choice assessments and to assess other dimensions of teaching, like classroom practice, that are difficult to capture through student assessments. Moreover, these teacher evaluation systems had the secondary goal to provide "relevant coaching, induction support, and/or professional development" (U.S. Department of Education, 2009, p. 9) to participating teachers. Papay (2012) made a similar call in describing the potential of both standards-based evaluations and value-added measures as potential tools for teacher professional development. He argued that effective teacher evaluation should provide "a clear understanding not only of [teachers'] current success or failure, but also of the practices they need to develop to become more successful with their students" (p. 138). Since this call, two experiments have tested whether teachers can leverage feedback from teacher evaluations to improve teaching and learning outcomes (Song et al., 2021; Steinberg & Sartain, 2015). Song et al. (2021) randomly assigned school districts to conduct up to four teacher observation cycles during the same school year with the goal to provide feedback to teachers about their teaching practice. Steinberg and Sartain (2015) studied the random roll-out of a new teacher evaluation system in Chicago public schools. Both studies found positive effects of receiving evaluative feedback on classroom practice (Song et al., 2021) and student learning outcomes (Steinberg & Sartain, 2015).

Other work has focused on non-experimental effects of teacher evaluation systems as a whole on teachers' development. Participating in a formal teacher evaluation system appears to improve student test scores outcomes between 10 and 25 percentage points (Dee & Wyckoff, 2015; Taylor & Tyler, 2012). However, these two studies are not able to fully identify the mechanisms behind these improvements. On the hand, Taylor and Tyler (2012) hypothesized that teachers participating in Cincinnati's teacher evaluation process learned new information about their performance, allowing them to adjust their teaching practice, leading to improvements in their productivity. On the other hand, Dee and Wyckoff (2015) studied the effects of the Washington, D.C.'s IMPACT teacher evaluation system and found that the threat of dismissal based on teacher evaluation cutoffs might lead to improvements in student test scores instead of the feedback associated with the evaluation process. Similarly, in focusing on the principal's role in teachers' evaluations and the feedback that follows formal teacher evaluations, Grissom and Loeb (2011) only found marginal evidence that principal instructional guidance during teacher evaluation is associated with increased student achievement measures and teacher satisfaction.

Overall, this work suggests that participation in the teacher evaluation process can lead to changes in teaching practice and improvements in student outcomes, with feedback on teachers' practice being the most likely mechanism behind these changes. This intuition is in line with experimental evidence that has found that increasing feedback based on objective and actionable measures of teaching leads to improvements in student test scores (Koedel et al., 2017).

## School Working Conditions

Better working conditions is the last dimension of teaching support in this chapter. This type of support includes school-level efforts to develop and nurture a better working conditions for teachers, by either implementing policies or structures to promote better collaboration among teachers (e.g., grade-level meetings, professional learning communities) or establishing better school leadership practices (Ladd, 2011; Moore Johnson, 2006; Moore Johnson et al., 2012).

Collaboration with colleagues is an important factor in the development of new teachers' practice. Qualitative studies on the transition between teacher preparation and teaching demonstrate the importance of that colleagues play in new teachers' practice development (Lortie, 1975; Zeichner & Tabachnick, 1981), showing the importance of collaboration and professional learning communities among in-service teachers (e.g., Horn, 2010; Little, 1987).

Learning how to work with colleagues and take an active part in the school's professional community is another feature of successful professional development programs for early career teachers. For example, Ronfeldt et al. (2015) studied the relationship between teacher collaboration in instructional teams and student achievement. They find that better achievement gains in both math and reading scores are associated with collaboration with colleagues. Moreover, teachers' scores appear to grow at faster rates when they report better collaboration with colleagues. Teacher instructional teams can provide one way for teachers to collaborate with each other. Both Jackson and Bruegmann (2009) and Sun et al. (2016) showed that individual teachers' instructional effectiveness relates to their peers' instructional effectiveness. Jackson and Bruegmann (2009) concluded that up to 20 percent of the variation in teacher effectiveness can be accounted for by their grade-level peers and that early career teachers are more impacted by their peers. Sun et al. (2016) studied teachers' transfers across schools and found "strong positive spillover effects associated with the introduction of peers who are more effective than the incumbent teacher himself or herself" when a transferring teacher was introduced to a new instructional team. This work highlights the importance of the instructional collaboration for early career teacher development.

Other studies on school improvement in Chicago public schools (Bryk & Schneider, 2002;

Bryk et al., 2009) and in Charlotte–Mecklenburg schools (Kraft & Papay, 2014) have found a connection between early career teachers' instructional improvement and their schools' professional environments. Specifically, teachers who experienced more supportive professional environments, including supports from school principals and other school climate indicators, improved at a faster rate and had better student outcomes than their colleagues that did not have the same experiences. These studies highlight the importance of the overall school working conditions, beyond collaboration with colleagues, that early career teachers experience in developing effective instructional strategies.

## The Current Work

As research into the effects of teaching supports for early career teachers has moved away from evaluations of single programs (Fletcher et al., 2008; Glazerman et al., 2010; Isenberg et al., 2009; Rockoff, 2008), studies that decompose teaching supports into types of supports that then focus on the relationships of these supports to teacher and student outcomes have become more common. For example, both Smith et al. (2018) and Auletto (2021) collect multidimensional measures for the teaching supports that early career teachers could receive during their first few years of teaching. I follow their example in this chapter by using rich, multidimensional measures for the types of teaching supports available to early career teachers.[1]

There are gaps in this prior work that this chapter addresses. In particular, the majority of prior work has relied on cross-sectional data about the teaching supports that early career teachers received. This is the case for every study that has used the SASS and the TFS data collections (Kang & Berliner, 2012; Ronfeldt & McQueen, 2017; Smith & Ingersoll, 2004) or ad hoc samples of teachers (Auletto, 2021). The main limitation of cross-sectional data is that it does not follow the same teachers over time. This could introduce bias in the estimates—especially for survey data—where response rates on surveys, unobserved differences between teachers, and missing outcome data could bias the results. In this chapter, I use longitudinal survey and outcome data. This allows me to address concerns with prior work and to use teacher fixed effects to account to unobserved differences between teachers.

Another way to address issues with cross-sectional data is to randomly assign teachers to receive different teaching supports. In fact, Glazerman et al. (2010) and Isenberg et al. (2009) have done so. However, they find mostly null effects for the effects of participation in a comprehensive induction program on most student-level outcomes. What is not clear from this work, though, is whether the contrast between the treatment and control conditions was big enough to observe effects of the intervention (Kang & Berliner, 2012) or whether they had enough power to detect

[1] I provide more detail on the development and validation of these measures in Chapter 4.

a small experimental effect on student outcomes. While my work is observational, it leverages a state-wide teacher survey program and includes the universe of respondents for whom I can calculate measures for teaching supports. To my knowledge, this is among the largest samples of early career teachers and allows me to observe even small relationships between teaching supports and instructional effectiveness outcomes.

Finally, Smith et al. (2018) is the other longitudinal study of the relationship between teaching supports for early career teachers and growth in instructional practice. While the authors collect multidimensional measures for teaching supports, they rely on scores on the Mathematics Quality of Instruction (MQI) rubric to assess instructional improvement in early career teachers. In this chapter, I similarly develop multidimensional measures of types of teaching supports, but I take a somewhat different approach to outcomes. I rely on instructional effectiveness outcomes from the Tennessee teacher evaluation process. This decision gives me the opportunity to study policy-relevant outcomes that matter to teachers, as early career teachers in Tennessee have to meet certain benchmarks on the outcome measures I use in this chapter to be eligible for tenure.

## Research Questions

The questions that guide this chapter are:

RQ1. What is the association between overall teaching supports and instructional effectiveness outcomes, as measured by observational ratings and student achievement gains?

RQ2. Which types of teaching supports are related to instructional effectiveness outcomes?

## Methods

### Sample

This chapter uses data from the Tennessee State Longitudinal Data System (SLDS). These data include teacher-level assignment information, teacher-student link data, teacher preparation data for teachers prepared in Tennessee educator preparation programs, and educator survey data from which I calculate measures of teaching supports.

The analytic samples depend on the main outcome of interest. The first sample contains all teachers who received at least one observation rating and TES respondents that have at least one teaching supports measure. The second sample contains all teachers who received at least one TVAAS score and have at least one teaching supports measure.

Table 2.1 reports summary statistics for these two samples. I can link observation ratings to at least one survey response for 58,819 unique teachers and to TVAAS for 36,116 teachers. Looking across the samples, the observation ratings sample has slightly fewer women, Whiter and more Hispanic/Latino teacher, less Black/African American teachers, and fewer mathematics, science, and elementary school teachers, and more special education teachers. This is in line with the intuition that TVAAS scores are available only for teachers in tested grade levels and subjects, which could affect both the teacher and assignment characteristics for each sample.

Covariates for my analyses include sets of teacher-level and school-level variables that could affect the teacher growth trajectories. Teacher-level covariates include basic demographic characteristics such as gender, race/ethnicity, and age, as well as teacher preparation information. I also included teacher assignment information to develop a measure of the kind of assignments. These covariates allow me to control for observed differences among teachers that could be correlated with early career growth. School-level covariates include student body composition, attendance rates, and average achievement.

I also include contextual workplace characteristics that could influence teacher growth rates. First, I calculate the fraction of teachers that stay at the same school year-by-year. Ronfeldt (2012) has shown that this stay-ratio measure is a signal for the working conditions of student teaching experiences. It is reasonable to assume that this stay-ratio might also predict the growth trajectories of teachers during their early careers, especially given prior evidence that working conditions in schools are related to better workforce outcomes (Moore Johnson et al., 2012).

Table 2.2 reports the differences between teachers who responded to the survey for whom I can calculate at least one measure of teaching support and teachers who did not respond to the survey. These results show that survey respondents are more likely to be instructionally effective (by 0.13 observation rubric points and 0.09 TVAAS standard deviation units), women (by 3.5 percent), White (by 7.7 percent), older (by 0.9 years), more experienced (by 0.7 years), endorsed in mathematics (by 1 percent) and elementary education (by 13.2 percent). These results raise some concerns about the external validity of my results, as my analytic sample is not representative of all teachers in the state (Joint test of significance $\chi^2(17) = 4,163.908$, $p < 0.001$). Caution then should be taken in interpreting the results and in extending them beyond the sample of respondents.

## Instructional Effectiveness Outcomes

Teacher effectiveness ratings on Tennessee's state-wide teacher evaluation system are the two main outcomes of interest for this project. These measures include observation ratings and Tennessee Value-Added Evaluation System (TVAAS) scores. Observation ratings are collected yearly for nearly all teachers in the state. Tennessee has developed an observation rubric—the Tennessee

Table 2.1: Sample Characteristics

| | Observation Ratings | | TVAAS | |
|---|---|---|---|---|
| | (1) Mean | (2) Std. Dev. | (3) Mean | (4) Std. Dev. |
| *Panel A: Direct Supports* | | | | |
| Teaching Supports | 0.019 | 0.833 | -0.014 | 0.856 |
| Mentoring | -0.010 | 0.536 | -0.036 | 0.563 |
| Professional Development | 0.002 | 0.781 | -0.022 | 0.825 |
| Evaluative Feedback | 0.013 | 0.757 | -0.015 | 0.780 |
| Working Conditions | -0.025 | 0.758 | -0.073 | 0.792 |
| *Panel B: Contextual Support* | | | | |
| Teaching Supports | 0.004 | 0.135 | 0.001 | 0.131 |
| Collaboration | 0.010 | 0.154 | 0.010 | 0.156 |
| Mentoring | 0.009 | 0.087 | 0.010 | 0.088 |
| Professional Development | -0.003 | 0.121 | 0.001 | 0.124 |
| Working Conditions | 0.016 | 0.119 | 0.012 | 0.114 |
| *Panel C: Covariates* | | | | |
| Evaluation Outcome | 4.000 | 0.517 | -0.017 | 0.838 |
| Women | 0.809 | 0.393 | 0.825 | 0.380 |
| Asian | 0.002 | 0.045 | 0.003 | 0.050 |
| Black/African American | 0.052 | 0.216 | 0.099 | 0.292 |
| Hispanic/Latino | 0.021 | 0.121 | 0.015 | 0.107 |
| Native American | 0.001 | 0.026 | 0.001 | 0.026 |
| Pacific Islander | 0.001 | 0.024 | 0.001 | 0.023 |
| White | 0.919 | 0.259 | 0.879 | 0.318 |
| Other | 0.003 | 0.051 | 0.003 | 0.048 |
| Age | 42.504 | 11.574 | 41.661 | 11.220 |
| Years of Experience | 11.505 | 9.976 | 10.938 | 9.499 |
| Education Level | 47.553 | 10.130 | 48.321 | 9.294 |
| Mathematics | 0.058 | 0.234 | 0.099 | 0.299 |
| Science | 0.053 | 0.224 | 0.077 | 0.267 |
| Special Education | 0.133 | 0.340 | 0.082 | 0.274 |
| Elementary | 0.770 | 0.421 | 0.840 | 0.366 |
| English Language Arts | 0.243 | 0.429 | 0.208 | 0.406 |
| Multiple | 1.257 | 0.712 | 1.307 | 0.672 |
| *N* | 58,819 | | 36,116 | |

Table 2.2: Comparison between Survey Respondents and Non-Respondents

| | (1) All | (2) Non-Resp. | (3) Resp. | (4) Diff | (5) Effect Size |
|---|---|---|---|---|---|
| *Panel A: Instructional Effectiveness Outcomes* | | | | | |
| Average Observation Score | 3.924 | 3.841 | 3.971 | 0.131 | 0.256*** |
| Average TVAAS | -0.066 | -0.131 | -0.045 | 0.086 | 0.113*** |
| *Panel B: Teacher Characteristics* | | | | | |
| Women | 0.796 | 0.773 | 0.808 | 0.035 | 0.087*** |
| Asian | 0.003 | 0.005 | 0.002 | -0.003 | 0.051 |
| Black/African American | 0.069 | 0.099 | 0.054 | -0.045 | 0.185 |
| Hispanic/Latino | 0.017 | 0.023 | 0.014 | -0.009 | 0.105 |
| Native American | 0.001 | 0.001 | 0.001 | 0.000 | 0.002 |
| Pacific Islander | 0.001 | 0.001 | 0.001 | 0.000 | 0.011 |
| White | 0.896 | 0.844 | 0.921 | 0.077 | 0.274*** |
| Other ERI | 0.012 | 0.023 | 0.006 | -0.017 | 0.187 |
| Age | 41.443 | 40.846 | 41.725 | 0.879 | 0.073*** |
| Teaching Experience | 10.524 | 10.045 | 10.759 | 0.714 | 0.070*** |
| *Panel C: Endorsements* | | | | | |
| Mathematics Endorsement | 0.056 | 0.049 | 0.059 | 0.010 | 0.045*** |
| Science Endorsement | 0.052 | 0.052 | 0.053 | 0.001 | 0.004 |
| Special Education Endorsement | 0.142 | 0.161 | 0.132 | -0.029 | 0.084 |
| Elementary Endorsement | 0.725 | 0.637 | 0.769 | 0.132 | 0.298*** |
| ELA Endorsement | 0.259 | 0.292 | 0.242 | -0.050 | 0.115 |
| Number of Endorsements | 1.234 | 1.191 | 1.255 | 0.064 | 0.088*** |

Note. Joint test of significance $\chi^2(17) = 4,163.908$, $p < 0.001$.
* p < 0.05 ** p < 0.01 *** p < 0.001

Educator Acceleration Model (TEAM)—that is used in most districts[2] in the state. The TEAM rubric assesses teachers on four domains (i.e., instruction, environment, planning, and professionalism) using twenty-four indicators. The number of observations that each teacher receives every year depends on their previous individual level of overall effectiveness (LOE). For teachers without tenure (i.e., teachers who hold a practitioner license and have less than three years of experience) the minimum number of observations is three for the instruction domain and two for the planning and environment domains. For teachers who have achieved tenure (i.e., hold a professional license and have at least three years of positive teacher evaluations), the number of teacher observations depends on their LOE; each step on the LOE scale reduces the number of observations. Regardless of tenure status, teachers who score in the highest level of effectiveness (i.e., level 5) are observed only once per year in each domain[3]. In addition to this number of minimum number of observations, at least half of the evaluation visits during a school year should be unannounced.

TVAAS scores are a type of teacher value-added model developed by SAS in cooperation with the Tennessee department of education. Conceptually, calculating a teacher value-added score involves two steps. In the first step, the model estimates student-level deviations from the expected growth that a student is predicted to make after controlling for school- and student-level covariates. In the second step, teacher scores are estimated as the average gains for students assigned to the same teacher. TVAAS scores are available for mathematics, English-language arts, science, and social studies for teachers in grades 2-8, end-of-course scores for Algebra I, Algebra II, English I, English II, English III, Biology, Chemistry, Geometry, Integrated Math I, Integrated Math II, Integrated Math III, and US History, ACT assessments in English, math, reading, and science, and other Advanced Placement (AP) assessments when available. Teacher scores are reported in student percentile scores (i.e., scores with a mean of 50 and standard deviation of 18.2). I convert these scores to teacher-standard deviation units (i.e., mean of 0 and standard deviation of 1) to allow for comparison of my results to more traditional value-added models using a conversion formula provided by SAS researchers.

## Measures of Teaching Supports Development

I develop measures for teaching supports from survey responses using psychometric methods. I started from reading questions included each administration of the First to the Top (FttT) survey and Tennessee Evaluation Survey (TES) and selected survey items that could measure teaching supports. I then matched these selected survey items year-to-year to identify which items were

[2] About one in ten teachers are assessed on non-TEAM rubrics. These teachers work in districts that have received a waiver from using the TEAM rubric after showing that their local observation rubrics are aligned with the TEAM rubric. The state department of education calculates equated observation scores for these teachers so that teacher evaluations are consistent throughout the state.

[3] Guidance on observation schedule is available here: https://perma.cc/9895-6ZTZ

administered in different survey years. This matching used a natural language processing algorithm that calculated the distance between the text of each item and matched items, even if the wording of the question changed, but its intent remained the same across survey instruments.

After selecting the survey items likely related to the four types of teaching supports I was interested in, I used psychometric methods to develop and validate measures for these four latent constructs. I first randomly split the survey respondents into two samples. I used the first sample for item selection and the second sample for latent construct validation. The use of two separate samples (rather than a single sample for both steps) reduces the risk of over-fitting a factor structure to a specific sample of respondents. In more detail, the first step involved an exploratory factor analyses (EFA). Through these steps, I selected survey items that loaded on the same measures of teaching supports and eliminated items that either did not strongly load on my construct of interest or loaded on more than one construct. The second step involved confirmatory factor analyses (CFA) on the second respondents' sample to validate the final factor structures for each teaching support measure form the EFAs and to estimate goodness of fit indices for each teaching support measures. All four measures displayed fair to excellent psychometric properties, indicating that these four measures are suitable for further use. Chapter 4 provides more details about the development and validation of these measures.

To reduce the dimensionality of the teaching support measures, I also developed a combined instructional measure as the principal component of the four kinds of teaching supports. However, this combined measure of teaching supports did not display the same psychometric properties as the individual measures of teaching supports. In other words, this combined measure of teaching support only roughly captures the variation in the four underlying types of teaching supports that teachers report experiencing. Nevertheless, I find this combined teaching support measure to be a useful summary of the overall teaching supports that teachers experience with the understanding that some details about teaching supports are lost in this combined measure.

## Measures of Teaching Supports Types

### Professional Learning & Development

Professional learning & development measure the extent to which survey respondents find professional learning & development opportunities at their school site supportive of their teaching practice. Items loading on this factor include questions about specific professional learning & development opportunities available to teachers and their connection to instructional improvement (e.g., "I receive specific professional learning suggestions that are tailored to my needs" or "In general, the professional development I have received this year has led to improvements in my teaching") and items about the role of school leadership or other teaching support staff for professional growth (e.g., “Leadership support such as key information and guidance from school or district administrators” or “Access to staff with specific expertise such as instructional coaches within and/or outside of my school”).

### Mentoring

Mentoring supports include an evaluation of the extent to which teachers felt these activities impacted their classroom practices. Questions focused on multiple outcomes, from influences on student learning (e.g., “Focuses on the aspects of my work that will affect student learning Select one option”), formative feedback on areas of strength and potential growth (e.g., “Helps me to identify areas where my teaching is strong” or “Provides me with clear expectations for my teaching.”), or usefulness in informing teaching practice (e.g., “Helps me plan instruction and develop lesson plans”).

### Evaluative Feedback

Evaluative feedback supports focus on the role of the teacher evaluation system and process in teacher professional growth. These items focus on the impact of participation in yearly teaching evaluations on teachers’ practice. Items include general questions about the impact of the teacher evaluation process on teaching practice (e.g., “In general, the teacher evaluation process used in my school has led to improvements in my teaching”) to items focused on particular features of the teacher evaluation process, such as the potential for receiving feedback on teaching practice (e.g., “I receive very detailed feedback on my strengths and weaknesses through the evaluation process”) or the role of the teacher evaluation system in shaping school activities (e.g., “Indicators from the teacher observation rubric are often referenced in formal meetings where teaching is discussed”).

### Working Conditions

Working conditions measure a mix of teaching supports focused on the workload that teacher experience (e.g., “The individual planning time provided for teachers in my school is sufficient”), feelings of autonomy in shaping instruction (e.g., “Teachers have autonomy to make decisions about instruction e.g. pacing, materials, and pedagogy”), extent to which school leadership is supportive of teachers’ development (e.g., “School leadership provides useful feedback about my instructional practices”), or opportunities for career advancement at their school (e.g., “Teachers are encouraged to participate in school leadership roles”).

### Overall Teaching Supports

To streamline parts of the results section, I report estimates for a combined measure of teaching supports. This measure is the principal component of the four teaching support types that I discussed above. This combined teaching support measure, however, does not share the same psychometric properties as the four measures I used to construct it. That is, psychometric analysis suggests that this unidimensional measure is only a rough approximation of the four separate measures of teaching supports that I combined. For this reason, this measure is better interpreted as an overall teaching supports average emerging from the different multidimensional measures of specific types of teaching support.

## Teaching Experience

Years of teaching experience is the main variable that I use to identify early career teachers. Following state policy on teacher tenure, I define early career teachers as the group of teachers who are working towards a permanent teacher license and are not eligible for tenure in Tennessee. In 2011, Tennessee updated their teacher tenure policy as part of their Race to the Top application[4]. These new tenure guidelines include a minimum level of teaching experience (i.e., five years) and two consecutive years of instructional effectiveness ratings above expectations or significantly above expectations to be eligible for tenure. For this reason, I identify two groups of early career teachers. The first is pre-tenure teachers in their first three years of teaching. These teachers are not eligible for tenure yet and teacher evaluations from this period cannot be used in the teacher tenure process. The second is tenure-eligible teachers who have between four and six years of teaching experiences. These teachers become first eligible for tenure during their fifth year of teaching experience and their instructional effectiveness ratings from this period count towards tenure eligibility.

## Analytic Approach

I structure this chapter's analysis into three parts: (1) I discuss the analytic approach to estimate the growth trajectories for evaluation scores; (2) I introduce my preferred teacher fixed-effects models; and (3) I present robustness checks to the estimates from the preferred models.

### Growth-Curve Model

I model teacher growth on evaluation scores using a growth-curve model (Rabe-Hesketh & Skrondal, 2012). Conceptually, a growth-curve model is a regression model that includes time as

[4]See, for example, https://perma.cc/34QQ-PVA8 for an explanation of the teacher tenure process in Tennessee.

a covariate of interest and accounts for differences among individuals in some way. This model uses the longitudinal nature of the evaluation data to estimate within-individual changes in evaluation scores and to calculate between-individual differences over variables of interest. Traditionally, these models involve, at a minimum, the inclusion of an individual-level random effect to account for unobserved differences among individuals. Effectively, these random effects allow each individual to have independent starting points at time zero or, more formally, allows the model to explicitly account for intercept heterogeneity.

In an equation, these four-level mixed effects models estimate

$$\begin{aligned} Y_{itsl} =& \beta_0 + \beta_1 \cdot Support_{itsl} + f\left(Exp_{itsl}\right) \cdot B_{itsl} + Support_{itsl} \times f\left(Exp_{itsl}\right) \cdot \Gamma_{itsl} \\ &+ \lambda_l + \sigma_{sl} + \tau_{tsl} + \epsilon_{itsl} \end{aligned} \tag{2.1}$$

where $Y_{itsl}$ is the instructional effectiveness outcome of interest (e.g., an observation rating or TVAAS score) $i$ for teacher $t$ in school $s$ and district $l$. $Support_{itsl}$ is the measure teaching supports estimated for teacher $t$. The coefficient $\beta_1$ estimates the average relationship between an increase in one standard deviation in $Support_{itsl}$ and changes in evaluation score $Y_{itsl}$. $f\left(Exp_{it}\right)$ is a function of years of teaching experience. I discuss how I model this functional form in more detail in the section below. The vector of coefficients $B_{itsl}$ estimate the average growth in evaluation score $Y_{itsl}$ across all teachers in the state for the experience function $f\left(Exp_{it}\right)$. The interaction term $Support_{itsl} \times f\left(Exp_{itsl}\right)$ allows for the relationship between $Support_{itsl}$ and $Y_{itsl}$ to be different along the teacher experience profile. For this reason, the vector of coefficients $\Gamma_{itsl}$ has the coefficients of interest in this model. These coefficients estimate whether reports of teaching supports have a differential impact on changes in the evaluation scores $Y_{itsl}$ for more or less experienced teachers.

The error terms $\lambda_l$, $\sigma_{sl}$, and $\tau_{tsl}$ are nested random error terms that estimate the extent to which the residual terms for evaluation scores $Y_{itsl}$ are correlated within the same school district $l$, the same school $s$, and the same teacher $t$. Each of these error terms are nested to reflect the natural clustering of teachers in schools, and schools in school districts. Finally, $\epsilon_{itsl}$ is the residual error term clustered at the school district level. These models are estimated using maximum likelihood and the residual terms are clustered at the school district level.

**Modeling Experience.** In my preferred models, I use a semi-parametric linear spline to model growth on evaluation scores. This spline has four parts: year zero through year two, year three through year five, year six through year ten, and more than ten years of experience. The first two parts of the linear spline are of interest in my analyses. The first term represents the pre-tenure early teacher career period. Current Tennessee law mandates a minimum probationary period of three years.[5] The second term of the time spline represents the tenure eligible early career period. I added a node at year five to keep the period measured by this term consistent with the first term of the spline (i.e., both terms measure three years of teaching experience). The rest of the spline has two additional terms divided at year ten, resulting in two additional career stages: mid-career and late career. I chose this node to keep my results consistent prior literature that found that the returns to teaching experience taper off after ten years of teaching experience (e.g., Papay and Kraft, 2015).

I test for the robustness of my results to my decision of modeling growth using a linear spline in two ways. First, I estimated a fully non-parametric approach that uses indicators for each year of experience below year ten to assess whether my results are sensitive to the linear functional form the spline imposes on the relationship between time and evaluation scores. In fact, Papay and Kraft (2015) conclude that this non-parametric approach is the best suited to calculate non-biased coefficients for the returns of experience on evaluation scores. On the other hand, this non-parametric approach makes interpretation of growth estimates difficult, as growth is the year-by-year difference in mean evaluation scores. This growth estimate is readily calculated in the spline model. Moreover, Bardelli et al. (2021) found that my preferred semi-parametric spline functional form performs as well as the non-parametric indicator form for early career teachers, suggesting that a fully non-parametric model might not be necessary in this case. Second, I estimated a fully parametric model that uses a polynomial expression to model teaching experience. This model estimates the effect of higher-degree curvatures on the returns of experience on evaluation scores. Following Kraft and Papay (2014), I used a fourth degree polynomial for this functional form. These alternative functional forms show that the semi-parametric linear spline performs qualitatively similar to both of them, suggesting that my preferred functional form is appropriate for my data.

### Identifying Assumptions and Threats to Identification

The identifying assumptions for these growth-curve models are similar to simple ordinary least-square (OLS) regression models. Specifically, multi-level models make the same assumptions as OLS models and add distributional assumptions for the random effect terms (Gelman & Hill, 2006; Snijders & Bosker, 2011). The key identifying assumption for these growth models is that the residual term and the random effects are independent of other regression covariates (i.e., exogeneity assumption). For this reason, this assumption is the main threat to identification for these growth models. That is, the presence of an omitted variable that correlates both with my predictor of interest (i.e., the measure of teaching supports) and either the residual term or the

[5] To facilitate the interpretation of the regression results, I start counting experience from 0 years instead of 1, so three years of teaching experience translate to year 2 in my time variable.

random effects could introduce bias in these estimates. I assess the extent to which this bias is present in these growth models in multiple ways.

Fixed teachers characteristics are a possible candidate for these omitted variables. In this case, estimates from the growth-curve models could be biased if there exist teacher characteristics that correlate with both teaching supports and instructional effectiveness outcomes. For example, it is possible that teachers who are more enterprising might both report higher levels of support because they are more likely to seek out such support and receive higher teacher observation ratings because they are seen as hard workers. Not adjusting my estimates for this possible kind of effect could introduce bias in my estimates in a direction that is difficult to assess in advance. Including a fixed effect for each teacher could address this kind of bias at the cost of requiring more than one observation per teacher and by introducing a new assumption of additivity between the fixed effect and the outcome of interest. In this case, identification will rely on year-to-year changes in teacher effectiveness outcomes and their correlations with teaching supports.

Similarly, it is also possible that school level characteristics could correlate with both the teaching supports and the instructional effectiveness outcomes. For example, a school could have an uncaring and strict principal who makes teachers feel less supported and also gives lower teacher evaluation scores. Again, this could lead to bias in the estimates. I address this concern in two different ways. I add controls for observed school-level characteristics to all regression models. These controls could account for omitted variables that correlate with both the control variable themselves, teaching supports, and the instructional effectiveness outcomes. A second way to address this concern would be through including school fixed effects. However, such models' identification would largely rely upon variation introduced by teachers moving between schools and on having enough movers each year to support these models. This requirement would further constrain the analytic sample. Therefore, I add these school fixed effects only in a specific case that I discuss below.

A final identifying assumption in the growth-curve models is that the predictors of interest are measured without measurement error. For example, bias can emerge if a teacher systematically over-reports the teaching supports they experienced and if their over-reports are correlated with their instructional effectiveness scores. At face value, the use of confirmatory factor analysis (CFA) to calculate the measures of teaching supports should partially address this concern, as CFA is better at accounting for measurement error than other data reduction methods. On the other hand, there might still be residual measurement error in measures of teaching supports that are not addressed by CFA.

I can assess the extent to which my estimates are sensitive to measurement error in two separate ways. First, assuming that teachers working at the same school receive similar levels of teaching supports, I can use teaching support averages of teaching supports reported by a teacher's colleagues working at the same school as an instrument for that teacher's reports. Intuitively, these models will identify the level of teaching supports that an individual teacher should have experienced given what their colleagues experienced. Second, assuming that systematic measurement error would affect scores from the same individual equally over time, I can use lags and leads in teaching support scores to assess the extent to which instructional effectiveness scores are correlated to previous or future reports of teaching supports. Any common effects across these different models—and especially the models with lead measures—would suggest the presence of individual level measurement error.

In the sections below, I describe in more detail how I address these threats to identification.

### Teacher Fixed Effects Models

While random effects are appropriate for a growth model, its coefficients might be biased when an omitted variable correlates with both the regressor of interest, the outcome, and the random effects. In other words, the random effects account for clustering and nesting of the error terms but perform poorly in the presence of omitted variables. For this reason, I replace the nested random error terms with teacher fixed effects. As a regression equation, I estimate

$$\begin{aligned} Y_{itsl} = \gamma_0 + \gamma_1 \cdot Support_{itsl} + f\left(Exp_{itsl}\right) \cdot \Delta_{itsl} + Support_{itsl} \times f\left(Exp_{itsl}\right) \cdot \Psi_{itsl} \\ + \tau_t + \epsilon_{itsl} \end{aligned} \tag{2.2}$$

where a teacher-fixed effects term $\tau_t$ replaces the random effects structure in the growth-curve models. This teacher fixed effect controls for unobserved fixed factors (i.e., constant over time) at the teacher level that impact growth on evaluation scores. These factors could include the teacher preparation experiences that teachers had, their major choice, or their beliefs about teaching. $\epsilon_{itsl}$ is the residual error term clustered at the school district level. The rest of the terms in the regression equation mirror the ones described above for the growth models. $\Psi_{itsl}$ remains the vector of coefficients of interest from this regression and its interpretation is the same as for $\Gamma_{itsl}$ from equation 2.1. This results in these fixed effects models leverage within-teacher, across-year variation in measures of teaching supports as the identifying source of variation in these models.

### Instrumenting Individual Reports with Peer Reports

One of the main challenges with working with survey-based data is assessing the extent to which survey-based measures capture individual-level experiences as separate from measurement errors and other sources of random error variation. This is a concern especially for the reports of teaching supports that early career teachers report receiving. Arguably, these less experienced teachers might not have participated in enough teaching support opportunities to accurately assess the quality of the teaching supports that they experienced. For this reason, I use peer reports of teaching support experiences[6] to instrument the individual teacher's reports. Results from these models provide a bounding exercise for the relationship between teaching supports and instructional effectiveness outcomes as these models identify the level of expected teaching supports given what other teachers at the same school reported receiving.

These analyses use the following estimating equation:

$$\begin{aligned} Y_{itsl} =& \gamma_0 + \gamma_1 \cdot \widehat{Supp}_{itsl} + f\left(Exp_{itsl}\right) \cdot \Delta_{itsl} + \sigma_s \tau_t + \phi itsl \\ Supp_{itsl} =& \delta_0 + \delta_1 \cdot Peer_{itsl} + \rho_s + \nu_t + \xi_{itsl} \end{aligned} \tag{2.3}$$

where $\widehat{Supp}_{itsl}$ is the teaching support score for teacher $i$ instrumented using their peer teaching support score $Peer_{itsl}$. All other coefficients are analogue to the ones included in Equations 2.1 and 2.2, except for the addition of a school fixed effects $\sigma_s$ and $\rho_s$. These models are estimated using two-stage least square estimation with fixed effects. The standard errors $\phi_{itsl}$ and $\xi_{itsl}$ are clustered at the school district level.

Conceptually, these instrumental variable regressions identify the level of teaching supports that individual teachers should have received using the reports of their colleagues working at the same school. The school fixed effects in these models further account for differences across schools in teacher evaluation policies and identifies the relationship between teaching support and changes in instructional effectiveness outcomes for teachers who switch schools. Finally, these instrumental variable regressions should partially address concerns of endogeneity between teaching supports and teacher observation scores by removing within-teacher variation in their teaching supports which, arguably, could be biased. For example, two teachers receiving the same kind and quality of teaching support might report it differently on the yearly survey only because the first teacher knows that they are receiving higher teacher evaluations than their peer. These instrumental variable regressions should address this concern. Additionally, it is also possible that teachers are selected to receive less or more teaching supports. This selection in receiving teaching support might bias my estimates either upwards, if teachers who are more likely to grow are allocated more teaching support, or downwards, if teachers who are less likely to grow are allocated more teaching support. Again, the instrumental variable regression outline above should address this concern.

[6]Technically, these are leave-one-out school level averages of teaching supports where I calculate the average teaching support measure for all teachers working at the same school site as teacher $t$ leaving out teacher $t$'s scores from this average.

### Lags and Leads

Another challenge with working with longitudinal panel data is that survey responses could report on prior year level of teaching supports or anticipate promised future teaching supports. For this reason, I conduct a second robustness check that includes lags (i.e., from the previous instructional year) and leads (i.e., from the following instructional year) for the measures of teaching supports in my preferred models. These estimates assess the extent to which concurrent year instructional effectiveness outcomes are correlated with prior or future reports of teaching supports. Ideally, instructional effectiveness outcomes should be associated with same year teaching supports (i.e., main models) and could be associated with a lagged relationship with prior year teaching support. On the other hand, I do not expect that anticipation effects of receiving teaching supports (i.e., leads) would have a relationship with changes in instructional effectiveness outcomes.

Finally, the models with future measures of teaching supports serve as a falsification test, as it is reasonable to assume that future teaching supports should not affect prior year instructional effectiveness estimates. If these estimates find significant coefficients between future teaching supports and prior instructional effectiveness outcomes, they would suggest possible endogeneity in who receives more teaching supports (e.g., selection into receiving teaching supports on time-invariant teacher characteristics) or issues with how the measures of teaching supports were developed and validated.

### Separating Individual Teaching Support from Average Teaching Support to Peers

A final analysis in this chapter aims to tease apart teaching supports provided to individual teachers from the average teaching supports that other teachers working at the same school report receiving. These analyses address a substantive rather than a methodological concern with the results from my preferred models. Prior literature has suggested the need to separate direct teaching support to teachers from teaching supports merely present in their environments. For example, Cohen et al. (2003) argued that "studies of resource effects on student outcomes [often] fail to take account of how teachers adjust instruction in light of their judgments about students [leading to] likely misestimate and confound the resources used, those merely present, and their effects" (p. 133). This observation points towards the need to isolate teaching supports that are targeted directly to individual teachers from teaching supports that are made available to everyone at the same school, as individual teachers might not benefit from teaching supports that they do not receive directly. For this reason, I include the measures for the average level of peer teaching supports I described above in my preferred models. First, I replace individual teachers' scores with their peer scores. Second, I include individual and peer scores together into the same models. The coefficients for the peer measure of teaching supports assess the average level of support that other teachers working at the same school report. When added to models with individual measures of teaching supports, these coefficients intend to separate individual teaching support experiences from the experiences of their colleagues.

# Results

This results section is organized into four sections. First, I present estimates for the extent to which my factors have an overall impact on teachers' evaluation growth trajectories and their heterogeneity for early career teachers. To streamline this section, I start with reporting the results for the combined, single measure of teaching supports and move to presenting heterogeneity in the results for the different types of instructional measures I developed. I then present results of robustness checks using instrumental variable regressions, and lags and leads in measures of teaching supports. Finally, I report the results of models that separate direct supports and the larger school-wide teaching support context through peer scores.

## Instructional Growth and Teaching Supports

Table 2.3 reports the results of the association between measures of teaching supports and teacher observation ratings in columns 1 and 2, and for TVAAS in columns 3 and 4. For each set of outcomes I report the results of two separate models: (1) 4-level random effects models with observations scores nested within school year, then teachers, then schools, and finally districts in the odd columns and (2) teacher fixed-effects models.

Three main findings emerge from this analysis. First, the results for the main effect for teaching supports—that is, the average relationship across all teachers in the state—shows that improvement in observation ratings is related to receiving more teaching support, while improvement on TVAAS scores is unrelated to receiving more or less teaching support. Depending on specification, I find that observation ratings for teachers who report receiving one more standard deviation in teaching support grow, on average, between 0.031 and 0.049 observation rubric points or, to contextualize this finding, the growth that pre-tenure teacher experience in two to four months of teaching experience. On the other hand, TVAAS estimates are mostly zero and non-significant, suggesting that improvements in TVAAS scores are largely unrelated to this combined measure of teaching supports.

Second, the interaction terms between the experience splines and the measures of teaching supports show significant heterogeneity in this relationship over teachers' career stages. These coefficients can be interpreted as an additive relationship on top of the average direct effect of teaching supports for teachers at different stages of their careers. Results for observation ratings show that

Table 2.3: Teaching Supports and Instructional Effectiveness

| | **Observation Ratings** | | **TVAAS** | |
|---|---|---|---|---|
| | **(1)** **Random Effects** | **(2)** **Fixed Effects** | **(3)** **Random Effects** | **(4)** **Fixed Effects** |
| Teaching Supports | 0.031*** | 0.049*** | -0.016 | 0.000 |
| | (0.005) | (0.004) | (0.013) | (0.019) |
| Support×Pre-Tenure | 0.011*** | 0.009*** | 0.019* | 0.000 |
| | (0.003) | (0.003) | (0.009) | (0.012) |
| Support×Tenure Elig. | 0.003 | 0.001 | 0.002 | 0.001 |
| | (0.002) | (0.002) | (0.006) | (0.006) |
| Support×Mid-Career | -0.002 | -0.001 | -0.005 | -0.002 |
| | (0.001) | (0.001) | (0.003) | (0.004) |
| Support×Late Career | 0.001*** | 0.000 | 0.001 | -0.000 |
| | (0.000) | (0.000) | (0.001) | (0.001) |
| Constant | 3.549*** | 3.688*** | -0.260*** | -0.261 |
| | (0.016) | (0.036) | (0.019) | (0.145) |
| Experience Spline | Yes | Yes | Yes | Yes |
| *N* | 528,372 | 528,149 | 114,807 | 106,496 |

Notes. This table reports estimates for the relationship between the overall teaching support measure and instructional effectiveness outcomes—observation ratings and TVAAS scores. Teachers are divided into experience groups: Pre-tenure teachers have between zero and two years of experience; Tenure eligible teachers have between three and five years of experience; Mid-career teachers have between six and ten years of experience; Late career teachers have over ten years of experience.

Robust standard errors in parentheses. * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

teachers in the pre-tenure period (i.e., teachers in their first three years of teaching) benefit the most from receiving teaching support by growing about an extra 0.01 teacher observation rubric points when receiving teaching supports one standard deviation above the average. This is equivalent to about a 25 percent increase in teacher observation ratings for pre-tenure teachers who experience this increased level of teaching supports from the average teacher in the state.

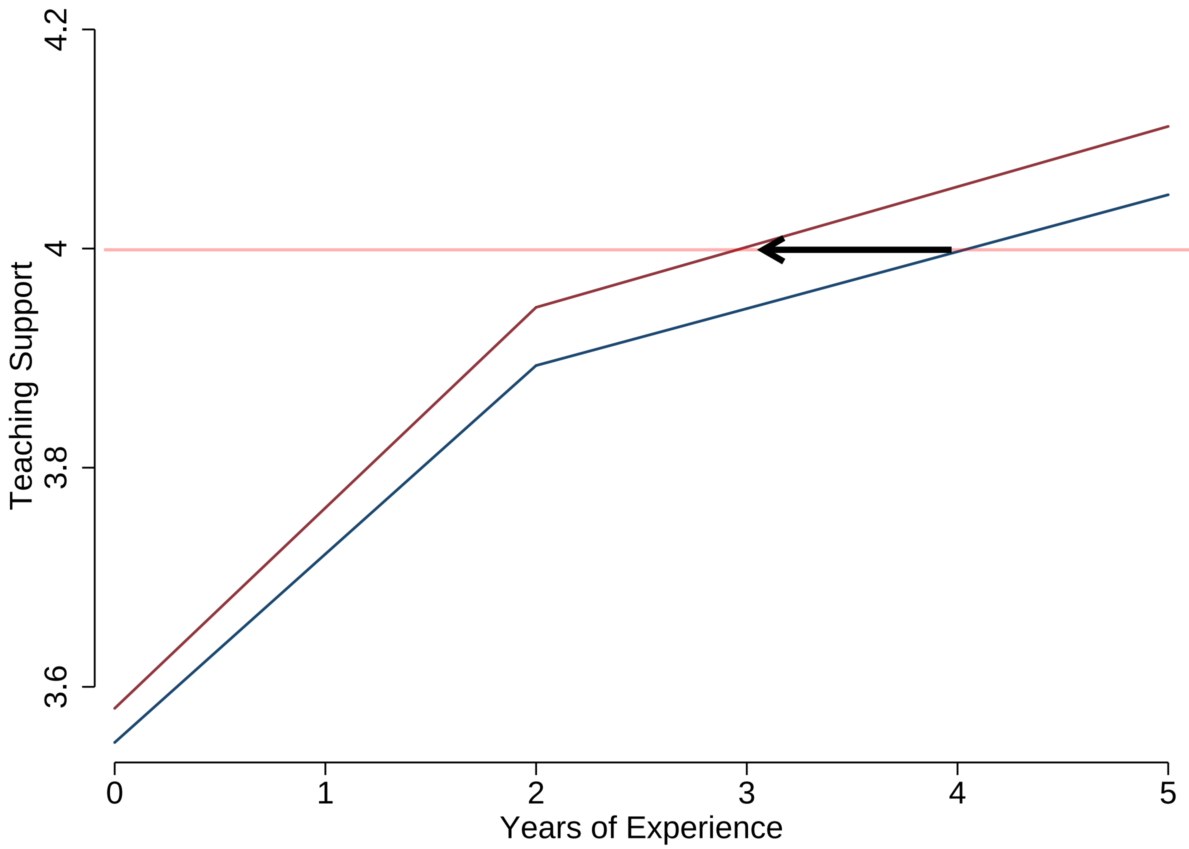
To better put this result into perspective, I calculate the expected observation ratings growth trajectories for teachers receiving average teaching supports and one standard deviation above the mean teaching supports over the first few years of teaching. Figure 2.2 visually reports the results of these analyses. The blue line represents the expected growth in observation ratings for a teacher experiencing average levels of teaching support; the red line represents the expected growth for a teacher experiencing one standard deviation greater than average teaching supports. Visually comparing the two lines, it is clear that there is a significant compounding process of receiving higher than average teaching supports on observation ratings, leading teachers who report receiving more teaching support each school year to reach an observation rating of four—or the required minimum for tenure—one year before their colleague that only received average teaching support during the same period.

Third, the results for TVAAS are mixed and varying by model specification. Results from the random effects models suggest that the benefits of receiving additional teaching supports are concentrated for pre-tenure teachers, with an improvement of 0.019 student standard deviations or about the growth in TVAAS scores we would expect with an additional five weeks of teaching experience for this group of teachers. Switching to teacher fixed-effects models, however, makes this relationship disappear. This suggests that the point estimates for random effects models might be identified from variation in teaching supports between teachers rather than variation in the same measures within a teacher. For example, improvement in TVAAS scores and teaching supports might be related to working in a more supportive school in general rather than receiving individual support.

## Heterogeneity by Type of Teaching Support

The results for the joint teaching support measure might hide some heterogeneity across the types of teaching supports that I used to build the combined measure of teaching support. I report the results for this heterogeneity analysis in Figure 2.3. Each coefficient reports the estimates for the association between each teaching supports types and growth in observation ratings, separate for pre-tenure teachers (in blue) and tenure eligible teachers (in red). Note that these coefficients are marginal expected values for each group or, in other words, the sum of the main effect for each measure and the interaction term between each measure and the experience group indicator.

Figure 2.2: Difference in Observation Ratings Growth



Appendix Tables A.1 and A.2 report the estimates for observation ratings and TVAAS scores for teacher fixed effects models.

Two main findings emerge from the analysis of observation ratings reported in Panel A. First, there is significant variation in the association between different types of teaching supports and growth in observation ratings. Second, there is additional variation in these associations between pre-tenure and tenure eligible teachers. In detail, both pre-tenure and tenure eligible teachers who report receiving more professional learning & development supports appear to grow at rates greater than the joint teaching support measure. Both mentoring supports and evaluative feedback supports show a different pattern. In this case, pre-tenure teachers who report receiving more supports along these types grow at faster rates than tenure eligible teachers reporting similar levels of support along these types. Finally, working conditions supports appear to have the lowest associations among the teaching support types I study, possibly stemming from the fact that these kinds of supports are more removed from classroom instruction than the other three.

I report the results for growth in TVAAS in Figure 2.3 panel B. These results show somewhat different results than the ones for observation ratings. In this case, I find little evidence that the relationships for each teaching supports type is significantly different from zero, as most confidence intervals cross the y-axis. Only mentoring appears to be significantly different from zero: Pre-tenure teachers who report receiving more mentoring appear to grow at a slower rate than peers that did not report receiving the same level of mentoring.

## Robustness Checks

### Instrumenting Individual Reports with Peer Reports

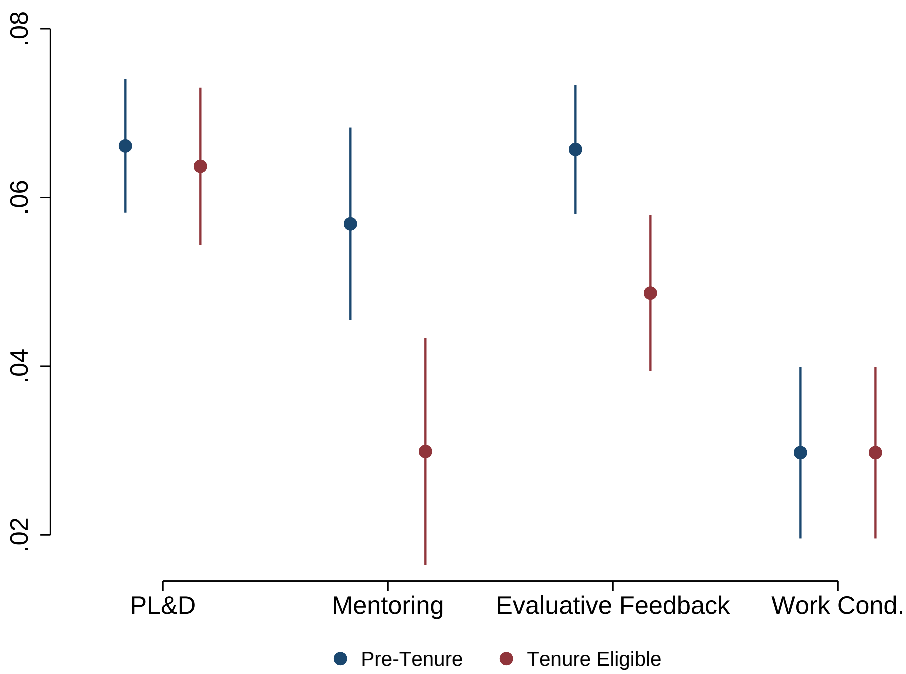
Table 2.4 reports the results of the instrumental variable regression of individual teaching support scores using peer measures of teaching supports for pre-tenure teachers. Column 1 reports the estimates for the second stage of the instrumental variable regression. Columns 2 and 3 provide additional information about these models, the reduced form estimates and the first stage estimates respectively.

Three main findings stand out in this table. First, the first stage appears to be appropriate for the instrumental variable approach. The F-statistic for the instrument is 1217.6, which is well above the customary cut-off values to identify weak instruments. Second, the reduced form is small in magnitude but still significant. Finally, the instrumental variable regression coefficient for teaching supports is 0.153, or about twice the size of the non-instrumented regression coefficient reported above.

This result suggests two takeaways from this robustness check. First, it suggests the concern that bias in individual reports of teaching support might be driving the results from my preferred models

Figure 2.3: Heterogeneity by teaching support Type

(a) Observation Ratings




(b) TVAAS

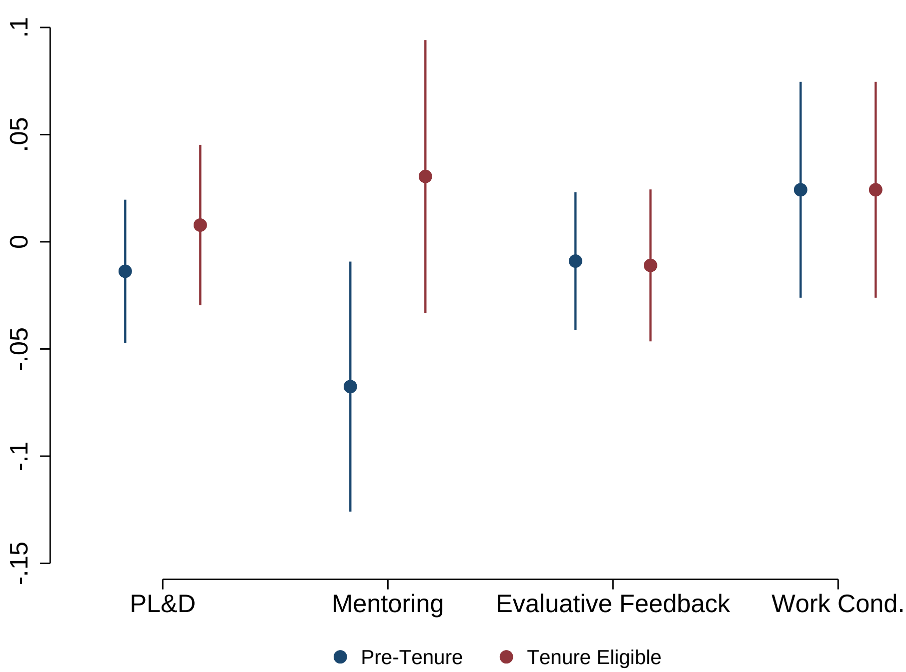


Table 2.4: Instrumental Variable Regression of Individual Reports using Peer Reports for Pre-Tenure Teachers

| | (1)<br>**IV** | (2)<br>**Reduced Form** | (3)<br>**First Stage** |
|---|---|---|---|
| Teaching Support | 0.153$^{*}$ | | |
| | (0.060) | | |
| Support to Peers | | 0.017$^{**}$ | 0.100$^{***}$ |
| | | (0.006) | (0.011) |
| Constant | | 3.704$^{***}$ | 0.491$^{***}$ |
| | | (0.040) | (0.037) |
| Experience Spline | Yes | Yes | Yes |
| Teacher Fixed Effects | Yes | Yes | Yes |
| School Fixed Effects | Yes | Yes | Yes |
| Instrument F-stat | | | 1217.616 |
| Observations | 112,281 | 112,281 | 121,317 |

Notes. Peer scores are calculated by averaging the measures of teaching supports for teachers working at the same school as the individual teacher, leaving out the individual teacher's scores.
Cluster robust standard errors in parentheses. $^{*}$ $p < 0.05$ $^{**}$ $p < 0.01$ $^{***}$ $p < 0.001$

is likely unwarranted, as the instrumental variable regression finds effects larger in magnitude than my preferred models. Second, the larger magnitude in this coefficient could be evidence of purposeful selection in who receives more or less teaching support within a school, with focus on teachers who are less likely to grow, leading to a deflation in the non-instrumented regression coefficients.

### Lags and Leads

Another issue in using survey panel data is the possibility that factor scores respond to lagged or leading (i.e., from the previous school year or the following one) teaching supports rather than teaching supports from the same school year. For example, teachers could report on teaching supports that they actually received the school year before or anticipate the results of upcoming teaching supports that are planned for an upcoming school year. For this reason, I reproduce my preferred teacher-fixed effects models but also include teaching support lags, leads, or both. This specification has the issue of restricting the estimation sample to teachers who have responded to the TES surveys either two or three times in a row.

Table 2.5 reports the results of these analyses. Again, I report estimates for the combined teaching support measure and for pre-tenure teachers. Column 1 reports the estimates for the

Table 2.5: Robustness to Lags and Leads

| | (1) Main Model | (2) Lags | (3) Leads | (4) Lags and Leads |
|---|---|---|---|---|
| Direct Supports | 0.061*** | 0.073*** | 0.056*** | 0.076*** |
| | (0.008) | (0.009) | (0.010) | (0.012) |
| Lag Support | | 0.027** | | 0.034*** |
| | | (0.009) | | (0.010) |
| Lead Support | | | -0.008 | 0.000 |
| | | | (0.011) | (0.011) |
| Constant | 3.851*** | 3.839*** | 3.870*** | 3.852*** |
| | (0.003) | (0.005) | (0.004) | (0.007) |
| Observations | 31,397 | 31,397 | 21,804 | 21,804 |

Notes. Lag support includes previous school year measures of teaching supports. Lead support includes subsequent school year measures of teaching supports.
Cluster robust standard errors in parentheses. * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

teachers who have responded to two surveys in a row; Column 2 adds the lagged measure of teaching supports from the previous academic year; Column 3 adds the lead for the same teaching support measure; Column 4 includes both the lag and lead measure.

Three main findings emerge from these robustness checks. First, the results from the same-year measures of teaching supports appear to be robust against the inclusion of both lags and leads in the same measure. Second, it appears that there are some left-over effects for lagged measures of teaching supports. These lagged measures appear to associate with future improvements in teacher observation ratings above and beyond same-year measure of teaching supports. Intuitively, these results could suggest either a delayed relationship of previous experiences with increased teaching supports or sustained improvements over time beyond the school year during which teachers experience increased teaching supports. Finally, I find little evidence of anticipation effects of measures of teaching supports, lining up with the expectation that future teaching support should flow back in time to previous year teacher observation ratings, playing the role of a simple falsification test. Moreover, these results provide evidence against the fact that teaching supports might be allocated to teachers who are more likely to grow, maybe because they are scoring lower on their teacher evaluations, as this process would require a positive association between future measures of instructional effectiveness and past instructional effectiveness outcomes.

Table 2.6: Robustness to Peer Teaching Support Scores for Pre-Tenure Teachers

| | (1) Individual | (2) Peer | (3) Both |
|---|---|---|---|
| Direct Supports | 0.067*** | | 0.064*** |
| | (0.004) | | (0.004) |
| Support to Peers | | 0.025*** | 0.017*** |
| | | (0.004) | (0.004) |
| Constant | 3.715*** | 3.731*** | 3.716*** |
| | (0.002) | (0.001) | (0.002) |
| Observations | 112,283 | 112,283 | 112,283 |

Notes. Peer scores are calculated by averaging the measures of teaching supports for teachers working at the same school as the individual teacher, leaving out the individual teacher's scores. Cluster robust standard errors in parentheses. * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

### Separating Individual Teaching Support from Average Teaching Support to Peers

Table 2.6 reports the results from these analyses for the combined measure of teaching supports. I report estimates for the combined measure of teaching support and for pre-tenure early career teachers in this table. Estimates for the four types of teaching supports and tenure eligible teachers are qualitatively identical to the ones reported in this table, and I omitted them for brevity. Column 1 reports the association between direct measures of teaching supports and teacher observation scores. Column 2 reports the association between peer measures of teaching supports and teacher observation scores. Column 3 reports the association between direct supports and teacher observation ratings after controlling for peer measures of teaching supports.

Two main findings emerge from this table. The first finding is that coefficients for direct supports are virtually identical between Column 1 and 3. This suggests that the relationship between reports of teaching supports and changes in instructional effectiveness outcomes are independent of the average level of teaching supports other teachers at the same school report receiving. Second, the average peer teaching support measure has a positive and significant relationship with changes in individual teachers' observation ratings, but this relationship decreases up to about a third when accounting for teachers' own measures of teaching supports. This finding could suggest that average levels of teaching supports experienced in schools could have an additive, independent relationship above and beyond the levels of teaching supports that individual teachers report.

# Discussion and Implications

Results from this study vary by outcome measure. Thus, I begin by discussing results and implications for observation ratings. After, I discuss results and implications for TVAAS.

## Observation Ratings

Teachers who report greater teaching supports have consistently greater growth in observation ratings. This relationship is both statistically significant and policy relevant, as teachers who report experiencing one standard deviation more teaching supports grow about twenty-five percent faster than their peers, leading them to become eligible for tenure up to a year earlier. These results are somewhat different from prior work that has found no relationship between organizational supports and improvement in instructional quality (Smith et al., 2018). One possible explanation for this difference could stem from differences in instructional effectiveness measures used. In the case of observation ratings, I find that these scores grow at a stead rate during the early career period. On the other hand, Smith et al. (2018) use scores from the Instructional Quality Assessment (IQA) rubric, which has been shows to be mostly stable during teachers' early careers (Desimone et al., 2016). The fact that my outcome of interest changes over time gives me the opportunity to relate these changes to the measures of teaching supports.

Moreover, I find some evidence that different types of teaching supports are associated with greater gains in observation ratings. Specifically, professional learning & development supports and evaluative feedback have stronger associations with growth than the other types of supports and working conditions have the lowest among them. These results are in line with my initial hypothesis that types of teaching supports more closely associated with instructional practice (e.g., professional learning & development) have the stronger relationship with observation ratings growth than other types that are more general (e.g., working conditions). However, all types of supports are related to positive growth in observation ratings. These results suggest that investing in teaching supports for early career teachers has promise, though investing specifically in professional leaning & development and evaluative feedback supports appear most promising. These findings suggest that the inclusion of professional learning & development supports in teacher induction programs could lead to improved observation ratings. Moreover, the association between evaluative feedback and growth in observation ratings could provide supporting evidence for the claims that teacher evaluation systems are a source of actionable feedback for teachers (Papay, 2012; U.S. Department of Education, 2009).

I find positive associations between teaching supports and growth in instructional effectiveness among all teachers, regardless of tenure status. However, the relationships are stronger among pre-tenure teachers. Specifically, the positive results for supports through the teacher evaluation system appear to be concentrated for teachers in their first three years of teaching, further highlighting the importance of teaching supports for teachers in their first years in the profession and possibly suggesting that these supports have diminishing returns as teachers become more experienced. These results could suggest that early career teachers benefit more from teaching supports and that we might expect this given that such supports would have the greatest benefits when first learning to teach, further bolstering the case for teaching induction programs or other formal supports for early career teachers.

Finally, it appears that experiencing better working conditions has a smaller in magnitude relationship with growth in observation ratings than the other teaching support types, suggesting that experiencing better working conditions might only indirectly relate to instructional improvement. In other words, it could be that better working conditions improve the environments where instruction happens, leading to indirect instructional improvement. Along the same line, it also appears that direct supports to individual teachers matter more than supports for their peers when considering improvements in these individual teachers' observation ratings. Together, these two results suggest that teaching supports that are directly provided to teachers, instead of teaching supports that target the school community rather than individual teachers, have the greatest relationship with observation scores improvement. These findings have policy relevant implications, as they suggest that policies and reforms regarding teaching supports are more likely to yield instructional improvements when these supports directly address individual teachers' needs as opposed to more global supports targeting school work environments generally. This is important as early career teachers tend to work in harder-to-staff schools with more challenging working environment than their more experienced peers (Bruno et al., 2019). Providing adequate teaching supports then holds promise for overcoming any challenges associated with these school placements.

## TVAAS

The results for TVAAS are more mixed. The multilevel models using nested random effects show a small positive relationship between teaching supports and TVAAS improvements; however, this relationship becomes null when using teacher fixed effects models. These results are largely in agreement with prior experimental work that has found no relationship between participating in a comprehensive teacher induction program and value-added to student test scores (Glazerman et al., 2010; Isenberg et al., 2009).

At the same time, the differences between the multilevel models and the teacher fixed effects models could suggest the presence of a possible source of omitted variable bias that is not accounted for in the random effects models, but that is controlled for by the teacher fixed effects models. For example, it could be possible that teachers who are more likely to improve their

TVAAS scores because they are self-motivated might also pursue more opportunities for teaching support. It is possible that the random effects models do not fully account for this third omitted variable, leading to somewhat inflated point estimates for the relationship between teaching supports and growth in TVAAS scores.

## Limitations and Directions for Future Work

The results in this chapter have several limitations. The use of survey-based measures of teaching supports limits my results in two main ways. First, my sample is limited to teachers who responded to the annual Tennessee Educator Survey (TES). As I find that the group of teachers who respond to the survey is significantly different from non-respondents on a variety of key characteristics, it is difficult to extend the results of my analyses beyond my analytic sample (i.e., survey respondents). Future work should address this limitation by collecting data from a more representative sample of teachers.

Second, the measures of teaching supports are limited by the questions included in the original TES instrument. While I was able to develop multiple measures for teaching supports, other possible kinds of teaching supports may be excluded from these analyses. Two candidates for these additional measures are salient. The first is a measure of leadership support, which has been shown in prior work to be associated with of early career teachers' development (Smith et al., 2018). The second measure is supports for the development of productive relationships with parents and the broader community surrounding the school. Bryk et al. (2009) suggested that this is an essential component for school improvement and teaching supports for relationship building could also lead to instructional effectiveness improvement for individual teachers. Future work should consider developing and validating measure for these additional dimensions of teaching supports and study their relationships with instructional improvement.

# CHAPTER 3

# Relations Among Early Career Teaching Supports, Professional Growth, and Employment Decisions

The increase in teacher shortages across school districts has made more pressing the development of innovations in how we recruit and initially prepare prospective teachers, especially in subjects that are considered high needs. However, less attention has been placed on what happens when these newly minted teachers enter the classroom and the conditions that help them remain in the profession. In this chapter, I study the relationship between a set of teaching supports new teachers experience and their relationship with mobility and attrition outcomes. The main goal of this work is to study whether prior evidence suggesting that teaching supports predict specific employment decisions (i.e., mobility and attrition) holds in a different labor market context, using new measures for teaching supports, and new methods to estimate the relationship between supports and employment decisions. Given that prior work has shown mobility and attrition to be associated with teachers' instructional effectiveness, this chapter also considers the role of instructional effectiveness in these relationships.

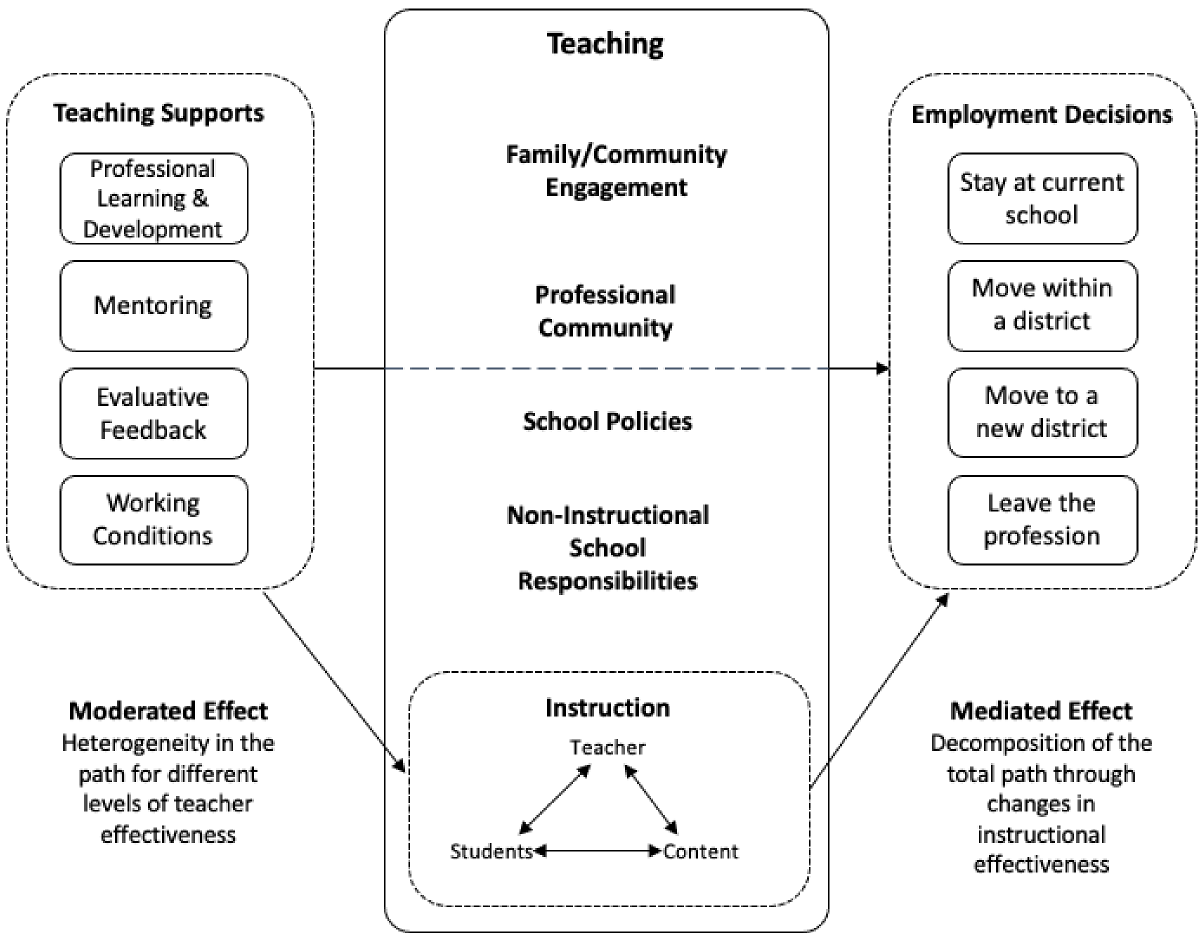
## Linking Teaching Supports to Employment Decisions

Figure 3.1 shows how I conceptualize the connection between teaching supports and employment decisions. I consider teaching as the full mediator between teaching supports and employment decisions, meaning that changes in teaching supports lead to changes in teaching and, in turn, these changes in teaching influence teachers' employment decisions.

### Teaching

Teaching sits at the center of my conceptualization. I define teaching as *the work that teachers do that is deliberately oriented to produce learning* (Cohen, 2011). Furthermore, I separate instructional activities, defined as the relationship between teacher, students, and content within

Figure 3.1: Teaching Supports, Instructional Effectiveness, and Employment Decisions



environments (i.e., the instructional triangle, Cohen et al., 2003), from non-instructional activities that teachers might engage in as part of their daily work, such as implementing school policies, taking up non-instructional responsibilities, engaging with the larger professional community at school, and building relationships with families and the community at large. This distinction is important because prior research suggests that instruction—as measured through instructional effectiveness—is associated with teacher employment decisions (Boyd et al., 2010; Goldhaber et al., 2010; Hanushek & Rivkin, 2010).

## Employment Decisions

Employment decisions are the main outcomes of interest in this chapter. Prior work on teacher employment has identified different types of teacher employment decisions: teacher mobility and teacher attrition. Teacher mobility refers to teachers moving between schools, either within the same school district or across different school districts. Teacher attrition refers to teachers leaving the profession altogether for another career path. While these two outcomes are somewhat different, they could have similar impacts[1] on schools, as a new teacher will be needed to replace the teacher that turned over (Ingersoll & Smith, 2004).

**Teacher Mobility.** In considering teacher mobility, it is helpful to separate within-districts and across-districts mobility. Teachers that move between schools but stay within the same school district usually do not experience changes in salary, as teachers working in the same district are usually on the same pay scale. This means that within-district mobility has a somewhat lower burden on teachers. Teachers who move within the same school district often do so to move to schools that teachers perceive as more desirable (e.g., prior studies have suggested this includes schools that enroll higher-income student; Greenberg and McCall, 1974), or in response to structural changes in the school system, like changes in student enrollment or in operations restructuring (Murnane, 1981). Mobility across school districts is somewhat more burdensome for teachers, as the new school districts might only recognize part of the teaching experience that teachers have and teachers have to potentially relocate to a new city and learn new district-wide policies and procedures.

[1]The direction of this impact is a point of contention in the literature. For example, Hanushek (2009) argued that purposeful deselection of low performing teachers could be used as a tool for school improvement. Others have pointed out that school administrators might not be able to consistently tie dismissals with instructional performance (Grissom & Bartanen, 2018), leading to detrimental effects of teacher attrition on schools (Bryk & Schneider, 2002; Ingersoll & Smith, 2004; Ronfeldt, 2012).

## Teaching Supports

### Professional Learning & Development

My measure for *professional learning & development* activities captures the extent, quality, and kind of professional growth opportunities provided to early career teachers within their school contexts. This measure includes formal and informal supports in schools that promote professional learning & development, including professional development programs and informal learning opportunities, such as collaboration with colleagues or having a supportive relationship with school leadership.

Although professional learning & development is one of the main teaching supports provided to early career teachers, relatively few studies have studied its relationship to teacher employment decisions (Nguyen et al., 2019). The consensus from these studies is that higher quality professional development predicts lower teacher attrition rates, both leaving the teacher profession and mobility to a different teaching position. For example, Allen and Sims (2017) studied participation in a professional development program for science departments in England. They found that attrition from departments decreased by 3 percentage points and attrition from the profession by 4 percentage points two years after opting for the professional development program. In a dissertation study, Erickson (2007) analyzed the 1999-2000 school and staffing survey (SASS) and the 2000-2001 teacher follow-up data to examine the relationship between participation and the intensity of professional development activities and teachers' employment decisions. The author found that participation in professional development activities decreases the likelihood of leaving the teaching profession, however the total number of hours spent in professional development was not related to employment decisions.

Focusing on early career teachers' specifically, teacher induction programs are arguably the most common form of professional development in which new teachers participate. The evidence between teacher retention and induction supports is mixed at best. Using nationally representative data from multiple SAAS and TFS data collections, Smith and Ingersoll (2004), Kang and Berliner (2012), and Ronfeldt and McQueen (2017) consistently found that participating in a more intensive teacher induction program[2] was associated with improved teacher retention.

Studies that have focused on single-district teacher induction programs found similar positive associations between program participation and teacher outcomes. For example, Matsko et al. (2007) analyzed the teacher induction experiences for about 1,700 teachers from Chicago Public Schools. The authors reported that over 80 percent of teachers in their sample reported receiving

[2] The SASS instrument included indicators for being assigned a mentor, participation in beginner's seminars, having collaboration or planning time with colleagues, receiving supportive communication from school leadership, reduction in teaching load or number of preparations, having a teacher's aide. Combined, these indicators are often referred to induction supports.

some sort of induction supports during their first years of teaching. They found that the intensity of the induction activities that teachers reported was positively associated with their overall satisfaction and intentions to stay in teaching and at the same school. Although they could not observe eventual teacher employment decisions, these results suggest that teachers who reported receiving more induction supports could be less likely to leave teaching or move to a different school.

Similarly, Auletto (2021) surveyed about 800 teachers from fifteen school districts in Michigan and asked about induction support experiences, job satisfaction, overall sense of support, and plans for the future. They found that different induction dimensions are correlated with different outcomes. For example, having a close colleague with whom early career teachers could collaborate was positively associated with retention intentions, while participating in professional development activities was positively associated with job satisfaction and overall sense of support.

Other work that had, perhaps, more complete samples of teachers from a single state or a more rigorous experimental design found mostly null effects of teacher induction programs on teacher retention. Wechsler et al. (2012) surveyed 2,670 teachers and mentors participating in one of the 39 induction programs active in Illinois in 2006 or 2008. While they found positive associations between participation in an induction program on teacher efficacy, they found no relationship between induction program participation and teacher employment decisions. They suggested two explanations for these results. First, they suggested that the economic downturn that Illinois experienced because of the 2009 Great Recession could have confounded any effect of teacher induction, as up to six percent of their participants were laid off during this period. Second, they found that school and district contexts were highly predictive of teacher retention, independent of participation in a teacher induction program. I will expand on this finding more in a later section.

Finally, Glazerman et al. (2010) and Isenberg et al. (2009) reported on the only large-scale randomized control trial testing whether a comprehensive teacher induction program impacted teacher retention. They found no relationship between participation in their comprehensive induction program that included sustained professional development activities and participants' retention. But they found that participation in this comprehensive teacher induction program was positively related to student outcomes for teachers who remained in the profession three years after the intervention.

### Mentoring

Mentoring, as part of new teacher induction, is arguably the most common teaching support provided to early career teachers, leading Ingersoll and Strong (2011) to conclude that "teacher mentoring programs have become a dominant form of teacher induction; indeed, the two terms are often used interchangeably" (p. 203). Despite its popularity, research, described in detail below, shows somewhat mixed results for the relationship between mentoring and teacher employment decisions. These contradictory findings suggest that more research, like the present study, is needed to understand how receiving mentoring relates to teacher retention.

Correlational studies have found mixed evidence of the relationship between early career teacher mentoring and teacher retention. Ingersoll and Strong (2011) reviewed the literature of the impacts of mentoring programs on early career retention. They found that, while mentoring was associated with some improvement on classroom instructional practices and student outcomes, teachers that reported participating in mentoring activities were not more likely to stay at their schools. In a related study, Ronfeldt and McQueen (2017) reanalyzed SASS data and linked the retention and mobility outcomes of early career teachers to the kinds of induction supports that they reported receiving. Their results suggest that mentoring might have a positive and small reduction in the probability of moving to a different school and a small and non-significant reduction in the probability of leaving teaching. From these results, it is not clear whether mentoring alone, or other school-based characteristics such as more supportive school leadership or supportive communication, are related to employment decisions, as mentoring was correlated with other teaching supports that early career teachers receive as part of teacher induction programs.

### Evaluative Feedback

Over the past ten years, teacher evaluation systems have become commonplace across the United States. One of the policy goals for implementing teacher evaluation systems was to provide objective and actionable feedback to teachers about their instructional performance, that can be used to inform professional development activities and teacher labor market decisions (U.S. Department of Education, 2009). Despite these intended goals, only a few studies have examined how and to what extent the feedback that teachers receive during their evaluations associate with teacher employment decisions. Most of these studies have focused on localized teacher labor markets and on introducing teacher evaluation reforms limited to those markets (Cullen et al., 2021; Dee & Wyckoff, 2015; Steinberg & Garrett, 2016). Taken together, these studies have found that introducing a formal teacher evaluation system is associated with an increase in turnover among lower performing teachers. The mechanism behind this increase is, however, unclear as these teacher evaluation systems included a mix of financial rewards for high-performing teachers and mandated dismissals for low-performing teachers (Cullen et al., 2021) or included a structured evaluative feedback system based on classroom observations and principal-teacher dialogue (Steinberg & Sartain, 2015).

In one of the few studies that has focused on a state-wide, low-stakes teacher evaluation system, Rodriguez et al. (2020) reported that the roll-out of such a system increased teacher turnover even without including punitive consequences for low-performing teachers. Moreover, the authors found that these attrition differences were concentrated in urban districts and low-performing schools, suggesting that low-performing teachers working in those settings might be more sensitive to teacher evaluation outcomes than their colleagues working in different schools and districts.

### Working Conditions

Teacher working conditions are the last type of teaching supports that I consider in this chapter. These measures of teaching supports summarize multiple facets of the work environment that early career teachers experience and that together shape their working conditions. These supports include an appropriate workload that affords enough planning time, having a collaborative work environment where collaboration with colleagues is supported and encouraged (e.g., instructional teams or professional learning communities), and having supportive school leadership. Overall, working in a school with better working conditions is associated with improved teacher quality and student outcomes (Moore Johnson, 2006; Moore Johnson et al., 2012), and a reduction in the intention to migrate schools (Geiger & Pivovarova, 2018; Ladd, 2011).

Focusing on specific aspects of working conditions, several studies have found that early career teachers who experience a more collaborative work environment are more likely to be retained. Multiple correlational studies, for example, using teacher survey data from the Baccalaureate and Beyond survey (BBS), the School and Staffing Survey (SASS), or the Teacher Follow-Up Survey (TFS) have found that induction practices that support collaboration among colleagues, such as common planning time allotted to working with other teachers at the school, predicted lower attrition rates among early career teachers (Henke et al., 2000; Kang & Berliner, 2012; Ronfeldt & McQueen, 2017; Smith & Ingersoll, 2004).

Other studies have found that early career teachers that reported having strong collegial relationship with their colleagues (Wechsler et al., 2012), experienced a better fit to their school community (Miller et al., 2020), or worked in schools with lower rates of teacher turnover (Sorensen & Ladd, 2020) tend to have better employment decisions. Bryk and Schneider (2002) stressed the importance of teacher collaboration by arguing that activities that develop trust among colleagues "such as openness to improvement, trust and respect, teachers having knowledge and skills, supportive leadership, and socialization [...] are more critical to the development of professional community than structural conditions" (p. 8). These results highlight the importance of the larger school professional community in developing and supporting early career teachers' employment decisions.

## The Role of Teaching Supports in Employment Decisions

The core theoretical assumption in this chapter is that teaching supports can influence employment decisions by affecting teachers' voluntary and avoidable turnover. Abelson (1987) outlines four different types of teacher turnover along employee-controlled and organization-controlled factors. Voluntary and avoidable turnover include employment decisions due to pay, working conditions or administration issues. In this case, a teacher decides to voluntarily leave their position to seek better employment somewhere else. These employment decisions could be avoidable insofar that the school or district could have improved salary or working conditions, for example, to address teachers' concerns. Involuntary and avoidable turnover includes dismissals, layoffs, or forced retirements initiated by school or district leadership that are outside the control of individual teachers. Voluntary and unavoidable turnover includes decisions due to, for example, family relocation or mid-career changes. In these instances, teachers elect to leave their positions for reasons outside the control of the school or district. Involuntary and unavoidable turnover include other reasons outside the control of the individual teacher or school, such as leave following sudden medical needs.

Voluntary and avoidable turnover is the focus of this paper, as teaching supports can directly influence this kind of employment decisions by changing teachers' employment decision-making process in two related ways. First, teaching supports could shape the teaching environments that teachers experience, leading to different employment decisions net of everything else. In this case, teaching supports directly influence employment decisions. For example, teachers that participated in peer mentoring programs might be less likely to want to leave their schools because they do not want to lose contact with their peer mentors. This relationship could be independent of whether peer mentoring had any effect on teachers' instructional practices, as it is often the case that early career teachers participating in teacher induction programs report feeling better supported even when their teaching practices do not change (Smith et al., 2018).

Second, teaching supports could lead to changes in instruction which, in turn, could impact the employment decisions. This relationship is an indirect one when employment decisions are based on changes in instructional effectiveness rather than experiences with teaching supports. Prior research found somewhat positive evidence that instructional effectiveness and teacher turnover are related (Boyd et al., 2010; Goldhaber et al., 2010; Hanushek & Rivkin, 2010). For example, Boyd et al. (2010) studied mobility and attrition patterns in New York City schools and found that more experienced and more effective teachers, as measured by their value-added contribution to student test scores, were less likely to both request a transfer and to leave teaching. Similarly, Goldhaber et al. (2010) studied the relationship between teacher value added measures and mobility in North Carolina. They found that more effective teachers were less like to move to different schools and to leave the North Carolina teacher workforce. On the other hand, Hanushek and Rivkin (2010) found no differences in teacher value-added to student test score measures between early career teachers who moved between schools, between districts, or left teaching and early career teachers who stayed in their assignments, suggesting that early career teacher attrition might not be associated with their instructional effectiveness. As a result, changes in instructional effectiveness could act as an additional mechanism through which teaching supports indirectly influence teacher employment decisions. Here, teachers who experience more teaching supports improve their instruction, which, in turn, makes it less likely for them to leave. This process could be an indirect effect if teaching supports only influence employment decisions through changes in instructional effectiveness; or it could be a mediator (or moderator, based on modeling decisions) if it only partially influences employment decisions through changes in instructional effectiveness.

## Contributions

This chapter intends to make several contributions to the literature on the relationships between teacher mobility and attrition and early career teaching supports. First, it contributes to our understanding of how teachers' employment decisions are related to teaching supports that teachers report experiencing. Unlike prior research that has focused on environmental factors correlating with teacher employment decisions (e.g., Nguyen et al., 2019), this work considers teaching supports available to early career teachers. The focus on the specific supports available to early career teachers and their connections to instructional practice has the potential to identify possible policy levers that school leaders and policymakers can use to improve the retention of early career teachers.

Second, most prior studies of early career teacher retention relied on either national survey data that only included a few questions about induction supports (Kang & Berliner, 2012; Ronfeldt & McQueen, 2017; Smith & Ingersoll, 2004) or to ad-hoc data collections that included small samples of teachers (Auletto, 2021; Smith et al., 2018). These two data collection designs have complementary weaknesses. Whereas large, nationally representative samples collect data from a large number participants across different settings, their survey-based measures of teaching supports only include few binary indicators for the types of activities in which teachers participated. Ad-hoc data collections, on the other hand, trade sample size and representatives for more detailed measures of teaching supports. In this chapter, I build upon the strengths of these prior studies by using statewide teacher survey data paired with rich administrative data to study how comprehensive measures of teaching supports relate to early career teacher employment decisions. This allows me to observe the universe of teachers responding to a state-wide survey and to use these responses to develop multidimensional measures of teaching supports.

Last, this work is, to my knowledge, the first to parse apart the direct effects of receiving teaching supports on retention from meditated effects through improvements in teachers' instructional effectiveness. No other prior work has simultaneously linked teaching supports, changes in instructional effectiveness, and employment decisions of early career teachers. The work in this chapter explores whether teacher instructional quality mediates the relationship between teaching supports and employment decisions. If I find teaching supports promote retention only indirectly (i.e., through improving instructional effectiveness) then this would suggest focusing on policies that promote instructional effectiveness since these would effectively also promote retention. This would also suggest that more research is needed on other teaching supports beyond those studied here to identify ones that promote retention directly.

## Research Questions

I ask two research questions to guide the work in this chapter:

RQ1. What is the relationship between different types of teaching supports and early career teachers' employment decisions?

RQ2. Does instructional effectiveness moderate the relationship between teaching supports and employment decisions?

## Methods

### Data

Data for this project comes from the Tennessee State Longitudinal Data System (SLDS). These data include teacher evaluation scores, student test scores, and employment decisions for teachers employed in Tennessee public schools. Access to statewide data is important when studying teacher employment decisions as teachers tend to move between districts within the same state. Moreover, teacher licensure is at the state level, so moving across state lines is often a costly decision especially during the early career years because most state only reciprocate permanent teacher licenses.

The main outcomes of interest are the year-by-year teacher employment status. First, I identify being employed as a teacher in a school if a teacher is reported as being evaluated as a teacher of record anytime during a school year. This includes receiving an observation rating or being linked to student test scores. The two processes that generate these evaluation scores are somewhat different. For observation ratings, school administrators report to the state the results of classroom observations through the TN-Compass online system. If a teacher is present in the teacher observation database, I assume that a school administrator has visited this teacher's classroom for at least sixty minutes during the school year.[3] For TVAAS scores, student test scores need to be linked to specific teachers. State policy allows for fractional assignments of students to teachers in case a teaching assignment is only part-time or to account for other teaching arraignments. These link data are generated and reported by local school districts and should reflect actual student placements sometime during the school year.

These two pieces of information allow me to identify who is plausibly a teacher during a given school year. These data also identify the school site where a teacher is employed. This allows me to generate the three variables of interest: being employed, leaving the profession the following year, or moving schools the following year. I identify being a leaver if a teacher is not reported as being employed the following school year. Similarly, I identify a teacher as a mover if a teacher stays employed but changes schools the following school year. This decision allows me to use school-level covariates contemporaneous with employment decisions in my regression models, as I am interested in modeling the relationships between the kinds of environments that teachers experienced the year that they decided, for example, to move between schools or leave the profession.

I also propose to use employment data from teacher employment records through the Personnel Information Reporting System (PIRS) dataset as a robustness check for my preferred way to identify teacher employment (using teacher evaluation data). The PIRS data are collected as part of administering the school employee retirement benefits. This dataset includes employment records of nearly the population of school public personnel in the state. From my prior work, identifying employment using PIRS is highly correlated with identifying employment using the evaluation data. However, this approach could over-identify teachers as leavers in cases they move to a school that does not report employment data to the state (e.g., charter schools) or under-identify leavers in cases when teacher moves to a non-instructional position on the teacher pay scale (e.g., instructional coach). Relying on teacher evaluation data, on the other hand, allows me to track employment decisions for teachers in instructional positions, as state law mandates that all teachers must be evaluated at least once during a school year. One limitation, though, of using evaluation data to identify employment patterns is that some districts and schools have missing evaluation data in some years.

## Sample

My sample includes 76,816 public school teachers in Tennessee for whom I can calculate measures of teaching supports. Table 3.1 reports the summary statistics for these teachers. The table

[3]State teacher evaluation policy mandates a minimum of sixty minutes of observation for each teacher that is evaluated using the TEAM rubric.

is divided into three sections by column; I first report statistics for all teachers, then I report separate statistics for pre-tenure teachers (i.e., teachers with fewer than three years of experience, *N* = 20,084) and tenure-eligible teachers (i.e., teachers between three and five years of teaching experience, *N*=16,157) separately. The panels separate the different variable categories: employment decisions, direct teaching supports, contextual teaching supports, and covariates.

Overall, I find that pre-tenure teachers are more mobile and tenure-eligible teachers are less mobile than the rest of the teachers. When considering both movement between and within school districts, pre-tenure teachers are more likely to both move between and within districts than their colleagues, while tenure-eligible teachers are less likely to move between districts and just about the same as the average in moving within districts. The probability of leaving the teaching profession is about the same for pre-tenure and all teachers, at about 9 percent, while it is reduced to about 7 percent for tenure-eligible teachers.

Moving on to direct teaching supports, pre-tenure teachers report receiving higher levels of teaching supports across all four measures when compared to the average teacher in the state. Tenure-eligible teachers follow a somewhat different pattern. These teachers report receiving more support through professional development, mentoring, and evaluative feedback and less support through collaboration than the average teacher in the state. I also observe no difference in the levels of contextual teaching supports reported by the different groups of teachers (see Panel C).

When comparing the teacher characteristics of these groups, I observe that pre-tenure teachers tend to score lower on observation ratings than their colleagues and that tenure-eligible teachers score right around the state average. This observation is in line with prior work that found that observation ratings growth at a faster rate during the early career and that this growth tapers off (Papay & Kraft, 2015). The only other difference in these characteristics is teacher age. For this variable, I observe that pre-tenure teachers and tenure-eligible teachers are younger than the average teacher in the state.

Table 3.2 compares the characteristics of survey respondents and non-respondents. It appears that the teachers who respond to the FttT/TES surveys and for whom I can calculate measures of teaching supports, are significantly different from teachers that do not respond to the surveys ($\chi^2(20) = 43,704.982$, $p < 0.001$). In more detail, teachers who respond to the survey are more likely to move to a different school (by about 0.6 percentage points), have higher average observation ratings (by 0.15 rubric points), be women (by about 6.4 percent) and White (by about 9.5 percent), have more teaching experience (by 0.39 years), and be endorsed in elementary education (by 7.6 percent). These results suggest that survey non-response is correlated with teaching characteristics, reducing the external validity of my results beyond my respondent sample.

Table 3.1: Sample Statistics

| | All | | Pre-Tenure | | Tenure Eligible | |
|---|---|---|---|---|---|---|
| | **(1)<br>Mean** | **(2)<br>Std. Dev.** | **(3)<br>Mean** | **(4)<br>Std. Dev.** | **(5)<br>Mean** | **(6)<br>Std. Dev.** |
| *Panel A: Employment Decisions* | | | | | | |
| Mover | 0.197 | 0.328 | 0.231 | 0.387 | 0.186 | 0.354 |
| Mover – School | 0.086 | 0.224 | 0.091 | 0.262 | 0.089 | 0.256 |
| Mover – District | 0.120 | 0.278 | 0.147 | 0.329 | 0.104 | 0.282 |
| Leaver | 0.094 | 0.254 | 0.095 | 0.278 | 0.069 | 0.238 |
| *Panel B: Teaching Supports* | | | | | | |
| Support | 0.022 | 0.828 | 0.205 | 0.911 | 0.038 | 0.895 |
| Professional Development | 0.018 | 0.778 | 0.187 | 0.859 | 0.031 | 0.842 |
| Mentoring | -0.011 | 0.540 | 0.062 | 0.572 | -0.001 | 0.575 |
| Evaluative Feedback | 0.011 | 0.754 | 0.179 | 0.826 | 0.030 | 0.817 |
| Working Conditions | -0.025 | 0.757 | 0.010 | 0.816 | -0.038 | 0.819 |
| *Panel C: Peer Teaching Supports* | | | | | | |
| Support | 0.004 | 0.107 | 0.003 | 0.139 | 0.009 | 0.144 |
| Professional Development | 0.004 | 0.097 | 0.003 | 0.126 | 0.007 | 0.131 |
| Mentoring | 0.009 | 0.085 | 0.004 | 0.087 | 0.009 | 0.088 |
| Evaluative Feedback | 0.013 | 0.093 | 0.012 | 0.122 | 0.019 | 0.127 |
| Working Conditions | 0.005 | 0.118 | -0.007 | 0.154 | 0.005 | 0.157 |
| *Panel D: Covariates* | | | | | | |
| Average Observation Rating | 4.023 | 0.518 | 3.753 | 0.512 | 4.034 | 0.498 |
| Women | 0.808 | 0.394 | 0.800 | 0.400 | 0.804 | 0.397 |
| Asian | 0.003 | 0.051 | 0.005 | 0.068 | 0.004 | 0.061 |
| Black | 0.099 | 0.290 | 0.082 | 0.269 | 0.086 | 0.276 |
| Hispanic/Latino | 0.022 | 0.122 | 0.026 | 0.148 | 0.023 | 0.138 |
| Native American | 0.001 | 0.028 | 0.001 | 0.035 | 0.001 | 0.032 |
| Pacific Islander | 0.001 | 0.024 | 0.001 | 0.034 | 0.001 | 0.029 |
| White | 0.869 | 0.323 | 0.871 | 0.325 | 0.881 | 0.316 |
| Other | 0.005 | 0.058 | 0.012 | 0.099 | 0.003 | 0.052 |
| Age | 42.852 | 11.664 | 33.347 | 9.560 | 36.060 | 9.022 |
| *Panel E: Endorsements* | | | | | | |
| Mathematics | 0.058 | 0.234 | 0.067 | 0.250 | 0.061 | 0.240 |
| Science | 0.053 | 0.225 | 0.059 | 0.235 | 0.047 | 0.212 |
| Special Education | 0.138 | 0.344 | 0.141 | 0.348 | 0.142 | 0.349 |
| Elementary | 0.763 | 0.425 | 0.684 | 0.465 | 0.724 | 0.447 |
| English Language Arts | 0.250 | 0.433 | 0.258 | 0.437 | 0.268 | 0.443 |
| Number of End. | 1.262 | 0.725 | 1.209 | 0.790 | 1.242 | 0.714 |
| *N* | 76,816 | | 20,084 | | 16,157 | |

Table 3.2: Differences Between Survey Respondents and Non-Respondents

| | **(1) All** | **(2) Non-Resp.** | **(3) Resp.** | **(4) Diff** | **(5) Effect Size** |
|---|---|---|---|---|---|
| *Panel A: Employment Decisions* | | | | | |
| Mover | 0.309 | 0.466 | 0.184 | -0.282 | 1.015 |
| Mover - School | 0.084 | 0.081 | 0.087 | 0.006 | 0.039*** |
| Mover - District | 0.229 | 0.390 | 0.102 | -0.289 | 1.063 |
| Leaver | 0.202 | 0.371 | 0.069 | -0.302 | 1.156 |
| *Panel B: Teacher Characteristics* | | | | | |
| Ave. Obs. Rating | 3.949 | 3.846 | 3.993 | 0.147 | 0.280*** |
| Women | 0.777 | 0.741 | 0.805 | 0.064 | 0.155*** |
| Asian | 0.004 | 0.005 | 0.003 | -0.002 | 0.032 |
| Black/Afr. Am. | 0.128 | 0.160 | 0.101 | -0.059 | 0.187 |
| Hispanic/Latino | 0.017 | 0.017 | 0.017 | 0.000 | 0.002 |
| Native American | 0.001 | 0.001 | 0.001 | 0.000 | 0.001 |
| Pacific Islander | 0.001 | 0.001 | 0.001 | 0.000 | 0.003 |
| White | 0.823 | 0.771 | 0.865 | 0.095 | 0.268*** |
| Other ERI | 0.024 | 0.042 | 0.010 | -0.032 | 0.238 |
| Age | 42.497 | 43.884 | 41.485 | -2.400 | 0.191 |
| Teaching Experience | 5.976 | 5.754 | 6.143 | 0.389 | 0.097*** |
| *Panel C: Endorsements* | | | | | |
| Mathematics | 0.058 | 0.058 | 0.057 | 0.000 | 0.002 |
| Science | 0.056 | 0.060 | 0.053 | -0.007 | 0.029 |
| Special Education | 0.141 | 0.147 | 0.136 | -0.011 | 0.031 |
| Elementary | 0.719 | 0.675 | 0.752 | 0.076 | 0.171*** |
| ELA | 0.267 | 0.292 | 0.248 | -0.044 | 0.100 |
| Number of End. | 1.240 | 1.232 | 1.246 | 0.014 | 0.019*** |

Note. Joint test of significance $\chi^2(20) = 43,704.982$, $p < 0.001$.
* p < 0.05 ** p < 0.01 *** p < 0.001

## Measures of Teaching Supports

This chapter focuses on four types of teaching supports. *Professional learning & development* supports include informal and formal learning activities aimed at developing early career teachers' instructional capacity. *Mentoring* supports include cycles of classroom observation and formative feedback carried out by an expert observer and instructional coach. Together, these two kinds of supports are often referred to as induction supports for early career teachers. *Evaluative feedback* supports aim to inform early career teachers about the teacher evaluation process in the state, supporting them through the evaluation cycle, showing them how to use information from their evaluations to improve classroom practices. Finally, *working conditions* measure features of the larger school community and environment that can support or constrain early career teacher learning and growth. These include teaching supports related to planning and preparing for instruction, promoting teacher collaboration, and more supportive school leadership.

I developed these measures of teaching supports using responses to the First to the Top (FttT) survey and the Tennessee Educator Survey (TES). These longitudinal survey programs started during the 2011-12 school year and continued ever since. The survey instrument itself is updated each year to add new survey items that reflect additional data collection interests, remove items that proved to be uninformative, or to rewrite items that proved to be confusing to respondents. For this reason, the set of items that I consider in these analyses is different for each survey administration year and I use common items between survey administrations to anchor measures year-to-year.

I developed the measures of teaching supports following psychometric best practices for latent factor development and validation. I first categorized and linked survey items across administrations to identify the set of survey items that could relate to my four measures of interest. I then used two randomly split samples of responses to select potential survey items that would load on the four constructs of interest in exploratory factor analyses (EFA) and to validate each measure using confirmatory factor analyses (CFA). The EFA stage identified a set of items for each survey administration year that could load together on the four measures of teaching supports that I was developing. CFA provides a robust set of tools to assess how well a group of items could measure a latent construct of interest and to estimate less biased estimates of the latent construct scores that account for both measurement error and random error (see, Chapter 4 for more details about this process). I conducted these two steps separately on two randomly selected sub-samples of survey respondents to reduce the risk of over-fitting the factor structure to the TES respondent sample. Overall, this analysis found that my four measures of teaching supports have fair to excellent psychometric properties, suggesting that these measures can capture enough covariation in the survey items to develop latent measures from them and justifying their use in this chapter.

I also combined these separate types of teaching supports in a single measure of general teaching supports intensity to better align my results to prior work (e.g., Ronfeldt and McQueen, 2017).

This unidimensional measure, however, had poor psychometric properties, suggesting that it might not fully capture the diversity of teaching supports that teachers receive. At the same time, this measure is helpful to further reduce the dimensionality of teaching supports to a single factor, with the caveat that this overall measure might miss some complexity in teaching supports that teachers experience.

## Analysis

I organize the analyses for this chapter into two parts. I first describe base multinomial probit models and linear probability models to estimate the relationship between teaching supports and employment decisions. I then describe multiple follow-up analyses to explore the robustness of the estimates from the base models and to explore the heterogeneity of the relationship between teaching supports and employment decisions.

### Modeling the Relationship Between Teaching Supports and Employment Decisions

**Multinomial Probit Models.** Multinomial probit models are an extension of probit models to a categorical variable with more than two outcomes. The intuition behind these models is that employment decisions are driven by an unobserved latent variable $Y^*$, usually interpreted as the combination of several factors in a single utility function. A teacher chooses between employment decisions by estimating the utility of each option and choosing the outcome with the highest utility. In my models, I assume that this utility function is a linear combination of the levels of teaching supports that teachers receive, their average observation ratings, and school site characteristics.

Formally, I assume that the probability of each employment outcome conditional on the regressors is normally distributed, that is

$$\Pr\left(Y_{it} = j | X\right) = \phi\left(\beta_0 + \beta_1 \cdot Supp_{it} + \beta_2 \cdot EG_{it} + X_{it} \times B_5 + \lambda_{it} + \epsilon_{it}\right) \tag{3.1}$$

where $\phi$ is the cumulative distribution function of the standard normal distribution, $Y_{it}$ is an indicator variable taking the value of 1 if teacher $i$ will experience the focus employment outcome $j$ the following school year $t$, $Supp_{it}$ is the measures of teaching supports that teacher $t$ reported receiving during school year $t$, $EG_{it}$ is a set of indicator variables identifying teacher $t$'s instructional effectiveness group during school year $t$, $X_{it}$ is a vector of employment school covariates including student body characteristics and teacher stay ratio, $\lambda_t$ is a teaching experience fixed effect for teacher $t$ in year $i$, and $\epsilon_{it}$ is the robust error term clustered at the school district level.

The main strength of using a multinomial probit model over separate LPMs is that the probability of each outcome is jointly estimated, and the standard errors are adjusted for the multiple outcomes. As it is difficult to directly interpret the probit regression coefficients, I report the conditional (partial) marginal adjusted predicted probability of $Y_{it} = j$ for a unit increase in measures of teaching supports when all other regressors are held at their means.

The inclusion of the experience fixed effects $\lambda_{it}$ turns these multinomial probit models into discrete-time hazard models and allows me to interpret the $\beta_1$ coefficient as the average change in the probability to observe a given outcome $j$ for all teachers. Moreover, these time fixed effects account for left and right censoring of observations that could arise from teachers entering and exiting my panel due to responding to the surveys at different stages of their careers and to teachers leaving the panel altogether if they stop teaching.

The set of indicator variables $EG_{it}$ divides teachers into three, mutually exclusive, instructional effectiveness groups. These variables account for the possible confounding effect of instructional effectiveness on the relationship between teaching supports and employment decisions. Intuitively, teaching supports could be allocated differently by level of instructional effectiveness. This could introduce bias in my estimates if, for example, teaching supports are prioritized to more instructionally effective teachers that principals want to retain at their school.

At the same time, some have raised concerns with using fixed effects in probit models (e.g., Arellano and Hahn, 2007; Fernández-Val, 2009). For this reason, I also model employment decisions using linear probability models.

**Linear Probability Models.** The linear probability model (LPM) estimates the change in probability of an employment outcome (i.e., moving to a different position or leaving teaching) conditional on changes in measures of teaching supports within the same teacher. Formally, I estimate the regression

$$Y_{iy} = \gamma_0 + \gamma_1 \cdot Supp_{iy} + \gamma_2 \cdot EG_{it} + X_{it} \times \Gamma_{+}\lambda_t + \xi_{it} \tag{3.2}$$

where the coefficients mirror the ones described for Equation 3.1. The coefficient of interest is $\beta_1$ that estimates the change in the percentage to observe outcome $Y$ for an increase in one standard deviation unit in the teaching supports variable $Supp_{iy}$. This coefficient is analogous to the conditional marginal adjusted predicted probabilities I report for the multinomial probit models.

### Follow-Up Analyses

I conduct several follow-up analyses to my main models to explore heterogeneity in the relationship between teaching supports reports and employment decisions and robustness checks using instrumental variable regressions.

**Heterogeneity by Career Stage.** This first set of follow-up analyses is inspired by the observation that early career teachers have different employment decisions than more experienced teachers. For example, early career teachers are more likely to leave teaching and to move between schools than more experienced teachers (see Table 3.1). For this reason, it could be possible that early career teachers might respond differently to teaching supports than more experienced teachers. In this set of analyses, I restrict my preferred models to teachers in the pre-tenure period (i.e., between zero and two years of experience) and in the tenure eligible period (i.e., between three and five years of experience) and estimate the same regression models described above on this restricted samples.

**Heterogeneity by Instructional Effectiveness.** Workforce outcomes could also be related to teachers' instructional effectiveness scores. The intuition for these models is that more (or less) instructionally effective teachers might be less (or more) likely to leave teaching. As a result, teachers at different ends of the instructional effectiveness distribution might respond differently to teaching supports. In these models, I create instructional effectiveness groups by dividing teachers into thirds within each year of experience. This process gives me three groups of instructional effectiveness within each year of teaching experience (i.e., lower third, mid-third, and upper third). I then interact the measures of teaching supports with these group indicators and report the relationship between teaching supports and employment decisions for each of these instructional effectiveness groups.

**Instrumental Variable Regression.** This follow-up analysis instruments an individual teacher's measures of teaching supports using the leave-one-out average of teaching supports by their colleagues working at that teacher's school. The intuition behind these analyses is that an individual teacher's measures of teaching supports might be endogenous with employment decisions. For example, it is possible that a teacher who plans to leave the teaching profession at the end of the school year would under-report the teaching supports they received. Another possible source of endogeneity in my preferred models is purposeful allocation of teaching supports to specific teachers. A principal might, for example, decide to better support a teacher that indicated a desire to move to a different school at the end of the school year in hopes that doing so might retain that teacher. Instrumenting an individual teacher's measures of teaching supports using their peers' reports helps address these kinds of endogeneity concerns with individual-level variation by identifying the level of teaching supports that a teacher should have experienced given what made available to everyone else working at the same school.

More formally, I estimate the following linear probability models using two-stage least squares

$$\begin{aligned} Y_{it} =& \beta_0 + \beta 1 \cdot \widehat{Supp}_{it} + \beta_2 \cdot EG_{it} + X_{it} \times B5 + \lambda_t + \sigma_{it} \\ Supp_{it} =& \gamma_0 + \gamma 1 \cdot PeerSupp_{it} + \gamma_2 \cdot EG_{it} + X_{it} \times \Gamma + \mu_t + \rho_{it} \end{aligned} \tag{3.3}$$

where $Y_{iy}$ is a workforce outcome (i.e., leaving, moving, moving school, or moving district) for teacher $i$ in school year $t$, $\widehat{Supp}_{it}$ is the instrumented measure of teaching supports for teacher $i$ in school year $t$, $PeerSupp_{it}$ is the instrument peer measures of teaching supports, and all other variables are analogues to what I described in Equations 3.1. The standard errors $\sigma_{it}$ and $\rho_{it}$ are clustered at the school district level.

**Individual-Reported and Peer-Reported Teaching Supports.** A complementary analysis to the instrumental variable regressions is to consider peer-reported teaching supports as potential measures of spill-over effects (on individual teacher's employment decisions) of teaching supports that are provided to all teachers working at the same school site. The intuition behind these models is that school environments where all teachers are better supported could influence an individual teacher's employment decisions above and beyond the teaching supports directly allocated to that teacher.

To explore whether these spill-over effects are present, I replace $Supp_{iy}$ with the leave-one-one school average in measures of teaching supports that teacher $t$'s colleagues working in the same school reported. These models estimate the relationship between contextual measures of teaching supports (i.e., teaching supports made available to everyone at the same school site) on teachers' employment decisions. I also include these contextual measures alongside individual measures in an effort to isolate the teaching supports individual teachers receive and the contextual supports that the rest of the school community receives.

**Mediation Analysis.** The last follow-up analysis is a mediation analysis that decomposes the relationship between teaching supports and employment decisions in a direct path and an indirect path through changes in teacher observation ratings. Intuitively, teaching supports could lead to higher observation scores which, in turn, lead to changes in employment decisions. This relationship could be above and beyond any direct effect between teaching supports and employment decisions.

Formally, I estimate this mediation analysis in the structural equation model

$$\begin{cases} Y_{it} =& \beta_0 + \beta_1 \cdot Supp_{it} + \beta_2 \cdot OR_{it} + \lambda_t + \sigma_{it} \\ OR_{it} =& \gamma_1 \cdot Supp_{it} + \mu_t + \rho_{it} \end{cases} \tag{3.4}$$

where $OR_{it}$ is the observation rating for teacher $t$ during year $t$ and the rest of the coefficients are analogues to the ones described for Equation 3.1. All models are estimated using maximum likelihood and confidence intervals are bootstrapped using 500 draws with replacement.

The coefficient $\beta_1$ estimates the direct effect between teaching supports and employment decisions. I calculate the indirect effect with $\beta_2 \cdot \gamma_1$ and the direct effect with $\beta_1 + \beta_2 \cdot \gamma_1$.

# Results

## Relationship between Teaching Supports and Employment Decisions

Table 3.3 reports the results for the analyses in this chapter. I further divide the table into two panels. In Panel A, I report the conditional (partial) marginal adjusted predicted probabilities from the multiple probit models regressing employment decisions on the measures of teaching supports. In Panel B, I report the coefficients of linear probability models. Each panel of the table reports the results for different teaching supports dimensions.

A scan of Column 1 reveals that an increase in *overall teaching supports* is associated with a decrease in the probability of all forms of teacher turnover (i.e., moving to a different school or district, or leaving teaching). For example, early career teachers who report experiencing one standard deviation above the mean in combined teaching supports are 0.8 percentage points (p.p.) less likely to leave teaching than their colleagues receiving average combined teaching supports. This is a reduction of about fifteen percent in the overall probability of leaving teaching. I find similar relationship sizes for moving to a different school or moving to a different school district.

Comparing the point estimates between the multinomial probit regressions and the linear probability models across the two panels shows that these two specifications have virtually identical estimates. This result suggests that either one of these modeling approaches can be used in further analyses. For the rest of this chapter, I report results from linear probability models for simplicity's sake.

Moreover, comparing the point estimates for each type of teaching supports (*learning & development, mentoring, evaluative feedback*, and *working conditions)* across Columns 2 through 5, shows that each measures of teaching supports is positively associated with a reduction in each form of teacher turnover, though the relationships tend to be strongest for *working conditions* and *mentoring*. Also, it appears that there is not much difference in the relationship between the different types of teaching supports and employment decisions. For this reason, I will focus on the combined measure of teaching supports moving forward.

Table 3.3: Relationship Between Teaching Supports and Employment Decisions

| | (1) SUPP | (2) PD&L | (3) MNT | (4) EF | (5) WC |
|---|---|---|---|---|---|
| *Panel A. Multinomial Probit Margins* | | | | | |
| Move School | -0.007*** | -0.007*** | -0.011*** | -0.008*** | -0.010*** |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Move District | -0.004*** | -0.004*** | -0.006*** | -0.004*** | -0.006*** |
| | (0.000) | (0.000) | (0.001) | (0.000) | (0.001) |
| Leave | -0.010*** | -0.010*** | -0.013*** | -0.010*** | -0.014*** |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Instr. Eff. Group | As Bal | As Bal | As Bal | As Bal | As Bal |
| School Controls | At Means | At Means | At Means | At Means | At Means |
| Observations | 110,169 | 109,328 | 76,945 | 109,328 | 76,945 |
| ADP of Moving School | 0.052 | 0.053 | 0.048 | 0.053 | 0.048 |
| ADP of Moving District | 0.018 | 0.018 | 0.018 | 0.018 | 0.018 |
| ADP of Leaving | 0.055 | 0.055 | 0.053 | 0.055 | 0.052 |
| *Panel B. Linear Probability Models* | | | | | |
| Move School | -0.009*** | -0.008*** | -0.012*** | -0.009*** | -0.012*** |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Move District | -0.016*** | -0.017*** | -0.022*** | -0.016*** | -0.021*** |
| | (0.001) | (0.001) | (0.002) | (0.001) | (0.001) |
| Leave | -0.011*** | -0.011*** | -0.014*** | -0.011*** | -0.014*** |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Instr. Eff. Group | Yes | Yes | Yes | Yes | Yes |
| School Controls | Yes | Yes | Yes | Yes | Yes |
| Observations | 110,169 | 109,328 | 76,945 | 109,328 | 76,945 |
| ADP of Moving School | 0.065 | 0.065 | 0.057 | 0.065 | 0.057 |
| ADP of Moving District | 0.087 | 0.087 | 0.084 | 0.087 | 0.084 |
| ADP of Leaving | 0.062 | 0.062 | 0.059 | 0.062 | 0.059 |

ADP Adjusted Predicted Probability
Cluster robust standard errors in parentheses. * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

## Heterogeneity by Career Stage

Table 3.4 reports the relationship between the combined measures of teaching supports and employment decisions for teachers at different stages of their careers. Comparing the coefficients reported in Columns 1 through 3 (i.e., comparing the average relationship for all teachers to the one for pre-tenure teachers or a tenure-eligible teachers), an increase in teaching supports relates to a decreased probability to move schools or districts, and to leave the profession regardless of career stage. Pre-tenure teachers that reported receiving one standard deviation more teaching supports than the average are 1.3 percentage points (p.p.) less likely to move to a different school, 3.9 p.p. less likely to move to a different district, and 2.5 p.p. less likely to leave the teaching profession than their peers that received an average level of teaching supports. These coefficients are about twice the size of the coefficients I observe for the average teacher in the state, suggesting that pre-tenure teachers are more responsive to receiving teaching supports than more experienced ones.

Table 3.4: Relationship Between Teaching Supports and Employment Decisions by Career Stage

| | **(1)**<br>**All** Teachers | **(2)**<br>**Pre-Tenure** | **(3)**<br>**Tenure Eligible** |
|---|---|---|---|
| Move School | -0.009*** | -0.013*** | -0.014*** |
| | (0.001) | (0.002) | (0.003) |
| Move District | -0.016*** | -0.039*** | -0.019*** |
| | (0.001) | (0.003) | (0.003) |
| Leave | -0.011*** | -0.025*** | -0.009*** |
| | (0.001) | (0.002) | (0.002) |
| Instructional Effectiveness Group | Yes | Yes | Yes |
| School Controls | Yes | Yes | Yes |
| Observations | 110169 | 17767 | 14545 |
| Adj. Predicted Prob. of Moving School | 0.065 | 0.080 | 0.079 |
| Adj. Predicted Prob. of Moving District | 0.087 | 0.135 | 0.094 |
| Adj. Predicted Prob. of Leaving | 0.062 | 0.078 | 0.057 |

Cluster robust standard errors in parentheses. * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

A somewhat similar pattern is also present for tenure eligible teachers. In this case, tenure eligible teachers appear to still respond to experiencing teaching supports in their decision to stay at the same school or district. However, I also observe that the coefficient between teaching supports and moving to a different district or leaving teaching is smaller in magnitude than for pre-tenure teachers and similar to the state average, suggesting a possible differential relationship between teaching supports and different employment decisions for tenure-eligible teachers. The point estimate for moving to a different school is virtually identical for tenure eligible and pre-tenure teachers.

Together, these results suggest that investment in teaching supports for all teachers, and especially early career teachers, is associated with an improvement in their retention, either by reducing their mobility between schools and districts or attrition out of the teaching profession. Moreover, teaching supports appear to increase the likelihood of an early career teacher staying at the same school, regardless of career stage.

In the next section, I explore the degree to which the relationship between teaching supports and retention might be mediated or moderated by teaching effectiveness. After all, instructional effectiveness improves the most during the early career years (Papay & Kraft, 2015), leading to possible heterogeneous relationships between teaching supports and employment decisions. In short, I next explore the extent to which the relationship between teaching supports and employment decisions is different for more or less instructionally effective teachers.

## Heterogeneity by Instructional Effectiveness Groups

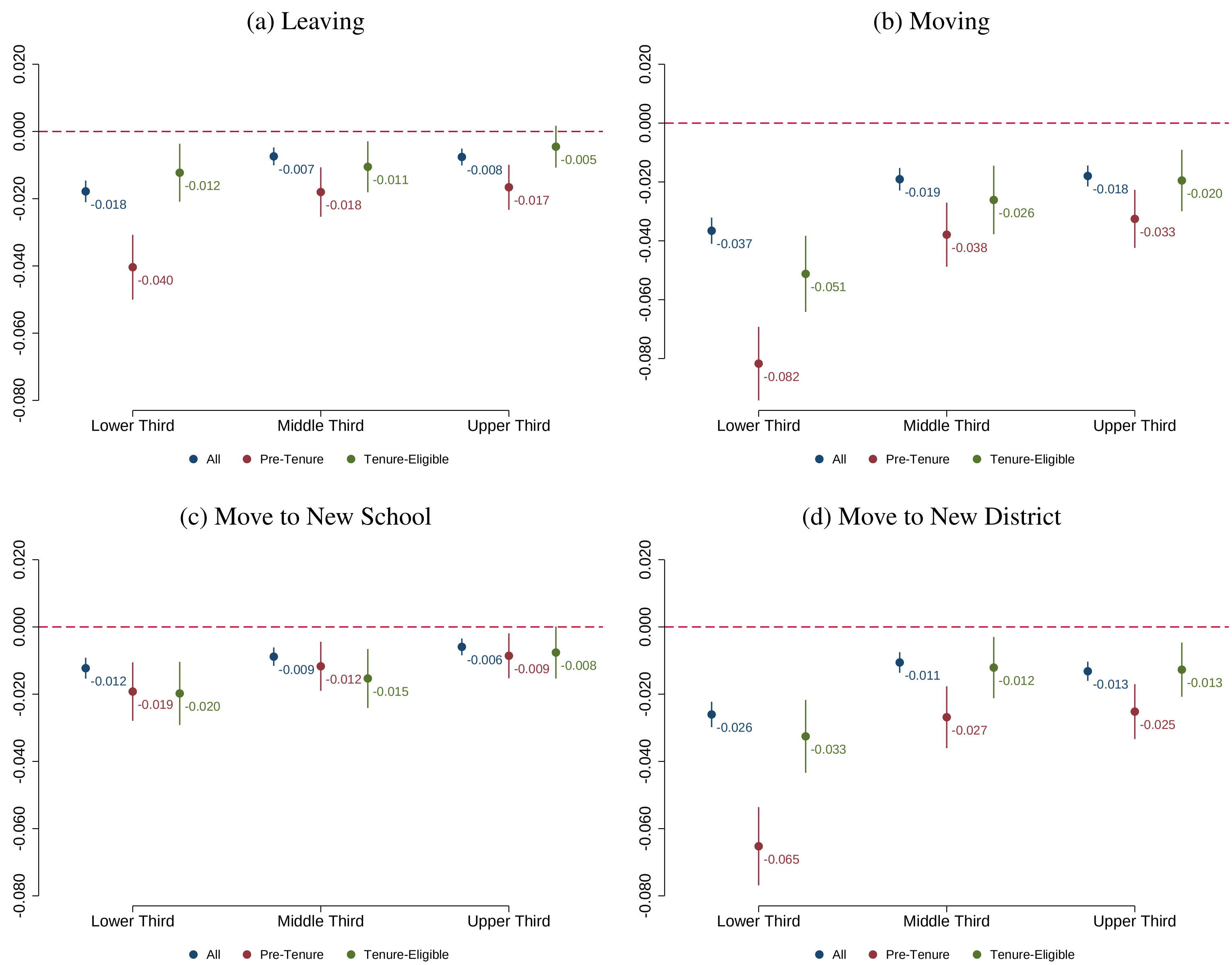
Figure 3.2 reports the marginal probabilities for each workforce outcome for teachers receiving an average observation rating (center), one standard deviation below (left) and above (right) the average. I further report estimates for all teachers in the state (in blue), pre-tenure teachers (in red), and tenure eligible teachers (in green). All coefficients report the effect of receiving one more standard deviation in teaching supports.

All coefficients appear to be lower on the left side of the graphs than the right side or, in other words, there is an increasing slope in the coefficients as observation ratings increase. This suggests that less instructionally effective teachers that report receiving more teaching supports are less likely to leave than their colleagues that receive less teaching supports. Moreover, these relationships appear to reduce in magnitude as teacher instructional effectiveness increases, suggesting that more effective teachers might be less responsive to receiving teaching supports.

Comparing the relationship of teaching supports to different employment decisions, it appears that receiving more teaching supports is related with a reduction in the probability of mobility to a different teaching position, while the probability of leaving teaching altogether might not the related to the level of teaching supports that teachers reported receiving. That is, the reduction in the probability of moving to a different school is about ten times the reduction in the probability of leaving.

Comparing the two mobility outcomes, it appears that retention within the same school district is improved by receiving teaching supports and that these improvements are concentrated for teachers scoring at the average or below in instructional effectiveness. Mobility between schools within the same district appear to still be related with a receiving teaching supports and this relationship

Figure 3.2: Heterogeneity by Instructional Effectiveness Groups



is not heterogeneous for teachers at different points of the instructional effectiveness distribution.

The difference in the relationship between same-school and same-district retention raises questions whether school-specific context has a relationship with improved employment decisions. For example, a teacher who receive teaching supports in a more supportive school environment might be less likely to leave a school than a colleague receiving a similar level of teaching supports in a less supportive school environment. I explore this relationship in the next section.

## Instrumenting Individual Reports with Peer Reports

Table 3.5 reports the result of instrumental variable (IV) regressions which instrument those supports reported by individual teachers using peer-reported teaching supports; I measure the latter using the leave-one-out average of teaching supports reported by teacher peers teaching at the same school. Column 1 reports the estimates from the second stage of the instrumental variable regression, Column 2 reports the reduced form estimates, and Column 3 reports the first stage estimates.

The IV regression estimates for both leaving and moving are qualitatively similar to the estimates from my preferred models reported in Table 3.3 Column 6. In more detail, the IV estimates are slightly smaller in magnitude by about ten percent, but these differences are not statistically significant. The IV estimate for leaving teaching is only marginally significant, perhaps indicating that these IV models are pushing the statistical power boundary for my sample of pre-tenure teachers. Finally, the estimates of the F-statistic for the first stage suggest that instrumenting those supports reported by individual teachers with their peers' reports provides a strong enough instrument for these IV models. Altogether, these results might partially address concerns in my preferred models with endogeneity between individual reports of teaching supports and employment decisions.

## Individual-Reported and Peer-Reported Teaching Supports

Table 3.6 reports the estimates comparing the relationship between individual and peer reports of teaching supports on teachers' workforce decisions. Reading the table from left to right, I report first the relationship for individual and peer teaching supports separately in Columns 1 and 2, and then the relationship from a regression model that includes both measures in Column 3. I finally report the coefficients from regressions that restrict the estimating sample to pre-tenure or tenure eligible teachers in Columns 4 and 5. Each panel of the table reports the coefficients for different employment decisions: leaving the teaching profession (Panel A) and moving to a different school (Panel B), and moving to a different district (Panel C).

Comparing the coefficients for individual teaching supports in Columns 1 and 3, it is clear that teachers who reported receiving greater teaching supports have better retention than their peers.

Table 3.5: Instrumental Variable Regression Estimates for Employment Decisions for Pre-Tenure Teachers

| | **(1)**<br>**IV** | **(2)**<br>**Reduced Form** | **(3)**<br>**First Stage** |
|---|---|---|---|
| *Panel A: Leave Teaching* | | | |
| Individual Support | -0.020$^{+}$ | | |
| | (0.011) | | |
| Peer Average Support | | -0.004$^{+}$ | 0.192$^{***}$ |
| | | (0.002) | (0.007) |
| Constant | | 0.117$^{***}$ | 0.114$^{***}$ |
| | | (0.004) | (0.013) |
| Instr. Effect. Group | Yes | Yes | Yes |
| School Controls | Yes | Yes | Yes |
| Adj. Predicted Prob. | 0.120 | | |
| F-stat | | | 727.789 |
| Observations | 17,767 | 17,767 | 17,767 |
| *Panel B: Move* | | | |
| Individual Support | -0.045$^{**}$ | | |
| | (0.016) | | |
| Peer Average Support | | -0.009$^{**}$ | 0.192$^{***}$ |
| | | (0.003) | (0.007) |
| Constant | | 0.303$^{***}$ | 0.114$^{***}$ |
| | | (0.006) | (0.013) |
| Instr. Effect. Group | Yes | Yes | Yes |
| School Controls | Yes | Yes | Yes |
| Adj. Predicted Prob. | 0.309 | | |
| F-stat | | | 727.789 |
| Observations | 17,767 | 17,767 | 17,767 |

Robust standard errors in parentheses. $^{+}$ $p < 0.10$ $^{*}$ $p < 0.05$ $^{**}$ $p < 0.01$ $^{***}$ $p < 0.001$

Table 3.6: Individual and Peer Supports on Employment Decisions

| | **(1) Direct** | **(2) Peer** | **(3) Both** | **(4) Pre-Tenure** | **(5) Tenure Eligible** |
|---|---|---|---|---|---|
| *Panel A: Leave* | | | | | |
| Individual Support | -0.011*** | | -0.010*** | -0.025*** | -0.009*** |
| | (0.001) | | (0.001) | (0.002) | (0.002) |
| Peer Average Support | | -0.003*** | -0.001 | 0.001 | -0.000 |
| | | (0.001) | (0.001) | (0.002) | (0.002) |
| Instr. Effect. Group | Yes | Yes | Yes | Yes | Yes |
| School Controls | Yes | Yes | Yes | Yes | Yes |
| Adj. Pred. Prob. | 0.062 | 0.062 | 0.062 | 0.078 | 0.057 |
| Observations | 110,169 | 110,166 | 110,166 | 17,767 | 14,545 |
| *Panel B: Move School* | | | | | |
| Individual Support | -0.009*** | | -0.008*** | -0.012*** | -0.013*** |
| | (0.001) | | (0.001) | (0.002) | (0.003) |
| Peer Average Support | | -0.005*** | -0.004*** | -0.003 | -0.005* |
| | | (0.001) | (0.001) | (0.002) | (0.002) |
| Instr. Effect. Group | Yes | Yes | Yes | Yes | Yes |
| School Controls | Yes | Yes | Yes | Yes | Yes |
| Adj. Pred. Prob. | 0.065 | 0.065 | 0.065 | 0.080 | 0.079 |
| Observations | 110,169 | 110,166 | 110,166 | 17,767 | 14,545 |
| *Panel C: Move District* | | | | | |
| Individual Support | -0.016*** | | -0.016*** | -0.039*** | -0.019*** |
| | (0.001) | | (0.001) | (0.003) | (0.003) |
| Peer Average Support | | -0.004*** | -0.001 | 0.002 | 0.001 |
| | | (0.001) | (0.001) | (0.003) | (0.003) |
| Instr. Effect. Group | Yes | Yes | Yes | Yes | Yes |
| School Controls | Yes | Yes | Yes | Yes | Yes |
| Adj. Predicted Prob. | 0.087 | 0.087 | 0.087 | 0.135 | 0.094 |
| Observations | 110,169 | 110,166 | 110,166 | 17,767 | 14,545 |

Cluster robust standard errors in parentheses. * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

These relationships are robust even after I include an estimate of the peer reported teaching supports. Moreover, it appears that working with peers who report higher levels of supports reduces the likelihood of leaving the profession, however this relationship is no longer significant when considering individual- and peer-reported teaching supports together in the same regression model.

Focusing on mobility outcomes (Panels B and C), the patterns for these analyses are similar to the ones described for leaving. The only exception is the coefficient for the relationship of peer reported instructional supports and moving schools. This coefficient remains significant when both individual and peer reported instructional supports are included in the same regression, likely indicating that the decision to stay at the same school is influenced by both individual level teaching supports and by spill-over effects of peer teaching supports.

### Mediation Analysis Through Observation Ratings

Table 3.7 reports the results of the mediation analyses for each workforce outcome for pretenure teachers. Overall, I find evidence that the relationship between teaching supports and workforce outcomes is significantly mediated by changes in observation ratings. That is, teaching supports correlate with improvements in observation ratings which, in turn, lead to a decrease in the probability of moving and leaving. At the same time, the direct relationship between teaching supports and workforce outcomes is only partially mediated (up to 14 percent for leaving and 17 percent for moving), indicating that improvements in observation ratings only partially explain the relationship between teaching supports and workforce outcomes. Thus, teaching supports appear to both directly and indirectly (through changes in instructional effectiveness) influence mobility and retention.

## Discussion

With the increased attention to teacher shortages, both from the predicted increased in teacher retirements in the coming years (Sutcher et al., 2016a, 2019) and from the pressure that the COVID-19 pandemic has put on the teacher workforce, there is an increase in interest in understanding how to better support teachers—and early career teachers in particular—to improve their employment decisions. Teacher turnover has been linked to both a decrease in student learning outcomes and instability of school working conditions (Hanushek et al., 2016; Henry & Redding, 2020; Ronfeldt et al., 2013; Sorensen & Ladd, 2020), as well as financial costs to districts that have to recruit and train new teachers (Watlington et al., 2010). It becomes important, then, to use historical teacher employment patterns to understand which teaching supports may lead to decreased teacher turnover. This chapter studies employment decisions for teachers working in Tennessee public

Table 3.7: Mediation Analysis

| | (1) Leaver | (2) Mover | (3) Move District | (4) Move School |
|---|---|---|---|---|
| Indirect Effect | -0.004*** | -0.010*** | -0.008*** | -0.002*** |
| | [-0.005,-0.003] | [-0.011,-0.009] | [-0.009,-0.007] | [-0.003,-0.002] |
| Direct Effect | -0.024*** | -0.048*** | -0.038*** | -0.011*** |
| | [-0.028,-0.020] | [-0.054,-0.043] | [-0.043,-0.033] | [-0.015,-0.008] |
| Total Effect | -0.028*** | -0.058*** | -0.046*** | -0.014*** |
| | [-0.033,-0.024] | [-0.064,-0.052] | [-0.051,-0.042] | [-0.017,-0.010] |
| Percent Indirect | 0.144*** | 0.172*** | 0.177*** | 0.166*** |
| | [0.112,0.176] | [0.147,0.196] | [0.149,0.205] | [0.106,0.226] |
| Percent Direct | 0.856*** | 0.828*** | 0.823*** | 0.834*** |
| | [0.824,0.888] | [0.804,0.853] | [0.795,0.851] | [0.774,0.894] |
| Observations | 24,011 | 24,011 | 24,011 | 24,011 |

Bootstrapped 95% confidence intervals in brackets. * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

schools and links these outcomes to measures of teacher reported measures of teaching supports. This longitudinal analysis finds several insights in the relationship between teaching supports and employment decisions.

My results suggest that teachers who report experiencing more teaching supports are less likely to turnover with, on average, a reduction of 0.077 percentage points (p.p.) for pre-tenure teachers from a baseline average turnover rate of 0.171 for pre-tenure teachers and a reduction of 0.042 p.p. on a baseline of 0.230 for tenure-eligible teachers. In other words, I find that experiencing more teaching supports reduces the turnover rate by 45% for pre-tenure teachers and by 18% for tenure-eligible teachers. These results are similar in magnitude to prior work that used the School and Staffing Survey to study teacher induction (Kang & Berliner, 2012; Ronfeldt & McQueen, 2017; Smith & Ingersoll, 2004).

I find some differences in how teaching supports could relate to employment decisions for pre-tenure and tenure-eligible teachers. My results suggest that about a third of the reduction in turnover is due to a decrease in the likelihood of pre-tenure teachers leaving the teaching profession altogether, one half is due to a reduction in moving to a different district, and the rest is due to a reduction in moving to a different school. These results are somewhat different for tenure-eligible teachers. In their case, a reduction in each of the three employment decisions roughly explains a third of the overall reduction in turnover. These results suggest that investment in teaching supports for early career teachers, and especially for teachers in their first few years of teaching, may be the most promising way to increase teacher retention. Moreover, the fact a reduction in moving between districts or leaving the teaching profession could explain most of the relationship between teaching supports and employment decisions suggests that programs aimed to support early career teachers might have the best chance for success when they are implemented by school districts or state-wide.

Prior work has highlighted the importance of more instructionally supportive schools in improving the retention of their teachers (Moore Johnson, 2006; Moore Johnson et al., 2012). My results seem to align with this intuition, as I find that same school retention is the only outcome that is associated with peer average scores. In other words, I find that the average teaching supports reported by teachers' colleagues at the same site has a strong association with a reduction in the probability of moving to a different school within the same district, arguably the easiest turnover event for a teacher to experience as this does not require relocation or commuting to a different school district or a change in profession. Moreover, these relationships appear to be independent of what individual teachers report receiving, suggesting an additive relationship between teaching supports provided to individual teachers and the overall school environment where they work.

Finally, I find that teachers' instructional effectiveness—as measured by observation ratings—appears to mediate only in part the relationship between teaching supports and employment decisions. I find that improvements in observation ratings that are associated with teaching supports only explain up to 15 percent of the total relationship between teaching supports and employment decisions. In other words, 85 percent of this relationship is not explained by changes in observation ratings. This finding suggests that instructional effectiveness only explains a limited portion of the overall association between teaching supports and retention. For example, if a teacher is contemplating leaving the profession, how good of a teacher they are only marginally predicts this decision. At the same time, teaching supports appear to strongly predict the same decision regardless of the association between teaching supports and instructional improvement.

Two observations are worth noting here about the mediation analysis. First, my estimate for the indirect effect of teaching supports through instructional improvement is not a causal estimate. Also, it likely underestimates the overall mediation effect of teaching supports through instructional improvement because any presence of measurement error or other issues around the reliability of the mediator tend to bias the mediation effect estimates towards zero. For this reason, my estimates should be interpreted as a possible floor for the mediation effect, meaning that if there is a mediated effect of teaching supports through instructional improvement, this effect should not be smaller than 15 percent of the total relationship between teaching supports and employment decisions. Second, although the remaining direct effect between teaching supports and employment decision is larger in magnitude than the mediation effect through instructional improvement, this latter coefficient remains significant in my models, which means that relationship between instructional supports and employment decisions is only partially mediated by improvements to instructional effectiveness.

## Contextualizing the Relationship Size between Teaching Supports and Employment Decisions

The number needed to treat[4] (NNT) coefficient is used in medical research to contextualize changes in percentages into a more concrete measure of the potential of therapeutic treatment by estimating the number of people who are needed to experience, on average, one extra desired outcome (Sackett & Haynes, 1997). In my case, the NNT estimates the number of teachers that are needed to experience one standard deviation increase in teaching supports to expect one more of them to be retained. Technically, the NNT is estimated by taking the reciprocal of the coefficients in Table 3.3.

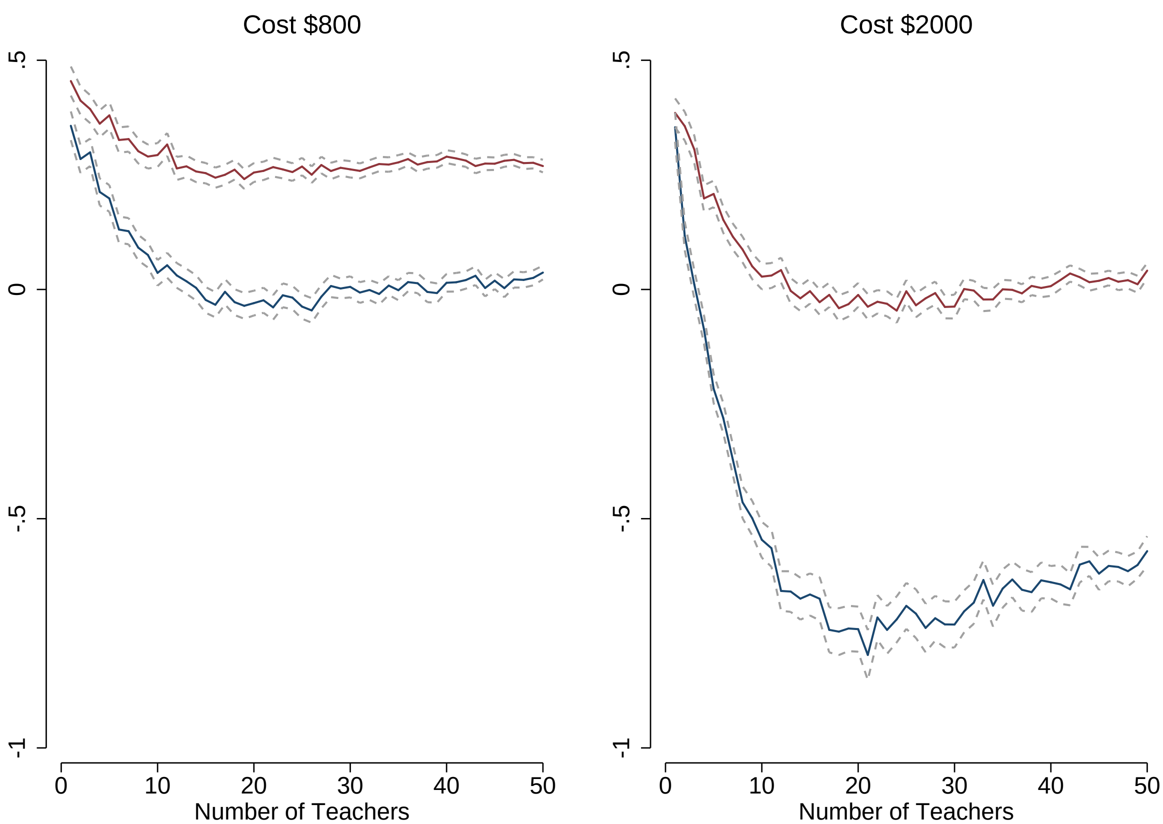
Calculating the NNTs for the relationship between teaching supports and employment decisions allows me to contextualize the relationship sizes in a policy-relevant measure. Focusing only on combined turnover reduction, the NNT suggests we would expect one more teacher not to turnover than expected when 13 teachers receive one standard deviation above the mean in the overall teaching supports. This finding is half the size to experimental results that linked a comprehensive teacher induction program to teacher retention (Glazerman et al., 2010; Isenberg et al., 2009), which suggested that the excess retention from the treatment condition (i.e., the improvement in retention because of treatment) is about six teachers or an NNT of about 25. These results help put into perspective the statistically significant reduction in teacher attrition that I observe for teachers who receive one standard deviation above the mean in teaching supports and suggest that most schools might not directly observe benefits of increased teacher retention only by providing teaching supports, as few schools have the number of early career teachers needed to observe a policy significant effect on teacher retention. On the other hand, these results suggest larger teacher induction programs that have the ability to enroll a larger number of teachers might be better suited to have the desired impact on teacher retention, highlighting the importance of district- or state-wide teaching supports policies and programs.

When looking at specific teacher mobility outcomes, my results further suggest that teaching supports alone may not result in a concrete improvement in specific employment decisions (i.e., same-school retention, same-district retention, or retention in teaching), as the number of teachers needed to experience the increase in teaching supports is somewhat large. For example, the NNTs for the overall measures of teaching supports for pre-tenure teacher suggest that 77 pre-tenure teachers are needed to experience one standard deviation above the mean in teaching supports to expect one more of them not to leave their school, 25 of them are needed for one more not to leave their district, and 40 of them are needed for one more of them not to leave teaching. These results

[4] I use the word "treat" and "treatment" here following its use in the medical field as in taking part in a therapy rather than assignment to a specific condition as part of a randomized control trial.

suggest that it is safe to assume that most schools would not be able to observe improvement in specific employment decisions by just providing teaching supports as they might not employ a large enough number of early career teachers to appreciatively observe declines in teacher mobility as a result of increases in teaching supports. However, school districts or states might employ enough early career teachers to make interventions targeting specific employment decisions viable. These results have also implications for researchers who want to assess the impacts of an intervention on specific employment decisions for early career teachers. These findings suggest that a large sample size, possibly in the thousands of participants, is needed to have enough power to detect differences in mobility or retention.

Figure 3.3: Simulation of the Percent Savings from Teaching Supports




Notes. Each line reports the results of simulations comparing the costs of providing teaching supports to an increasing number of teachers and the cost to replace districts that leave the profession after one year. The blue line reports the estimated cost/benefit ratio in for districts where the cost of replacement is low. The red line reports the estimated cost/benefit ratio in for districts where the cost of replacement is high. Positive y-axis values suggest that providing induction supports increases costs than just replacing teachers; Negative ratios suggest that providing induction supports reduces costs than just replacing teachers.

Another way to assess the policy relevance of providing teaching supports to early career teachers is to conduct a cost-benefit analysis. To this end, I run a simulation that assesses the trade-off between the costs of providing increased teaching supports to pre-tenure teachers and the reduction in turnover rates and related costs resulting from this increased support. The cost of teacher turnover is estimated to be between $10,000 and $25,000 depending on local labor market conditions (Watlington et al., 2010). Taking these estimates and scaling them up to 2022 dollars, I assume that the cost of replacing a teacher is between $12,500 and $31,000. I use these two estimates to run simulations for the possible financial costs and benefits associated with providing teaching supports to first-year teachers, assuming an average combined turnover rate for pre-tenure teachers to be 17.1% and that an increase in one standard deviation of teaching supports reduces this rate by 7.7%—both estimates stem from the rates I observe in my results.

Figure 3.3 reports the results of these simulations. I report the percent change in total costs in low-turnover cost districts in red and for high-turnover cost districts in blue. Positive values on the y-axis indicate cost saving associated with retention while negative values indicate cost increases. Each panel of the figure reports estimates associated with different costs for increasing the teaching supports for pre-tenure teachers to one standard deviation above the average. I find that the *breaking even* point of increasing teaching supports by one standard deviation above the average costs is $800 for low-turnover cost districts and $2,000 for high-turnover cost districts. That is, if providing the increased teaching supports costs less than $800 in low-turnover cost districts and $2,000 in high-turnover cost districts would result in an overall cost reduction of the combined cost of providing the increased teaching supports and replacing teachers that turnover when compared to the business as usual of just replacing teachers who turnover. These cost estimates, however, rise questions about whether it is possible to design a high quality teaching support program that fits within those cost constraints, as recent estimates suggest that traditional professional development workshops costs between $660 and $1,100 and that one hour of coaching costs $440 (Barrett & Pas, 2020).

My simulation reaches a stable point after about 40 teachers, suggesting that reduction in turnover costs balances out the increased costs of providing support to larger groups of teachers. Finally, I find some instability in the percent cost reduction when the group of teachers is less than twenty teachers. This could suggest that policies that aim to reduce costs associated with teacher turnover should apply to either large school districts that hire more than twenty teachers annually or regions (i.e., countries or intermediate education agencies) leading to pooling the potential cost reduction across smaller school districts.

## Limitations

This work has some remaining limitations that future work could address. First, I rely on teacher self-reports of teaching supports, collected through a state-wide survey of teachers that was not designed to measure these constructs specifically. While this data source gives me the opportunity to conduct one of the first longitudinal studies of the relationship between teaching supports and employment decisions, having better measures for teaching supports would strengthen this work. Two possible ways could improve upon my measurement model. Future work could develop measures for the kind and the extent of the teaching supports actually provided to teachers rather than relying on teacher self-reports. These instruments could collect, for example, a census of the professional development and activities that teachers participate in. Also, teacher surveys could be redesigned to directly ask about the teaching supports that they received and to better align with the underlying teaching supports dimensions.

Second, this work is a correlational study that relies on within-teacher variation in teaching supports to make inference on their relationship to employment decisions. Confounding factors that could correlate to both my measures of teaching supports and teachers' employment decisions could still bias my estimates in unpredictable ways. For this reason, the instruction supports I identify as promising should be tested under experimental or quasi-experimental conditions to address these concerns.

Third, as discussed above, this work is a secondary data analysis of a survey instrument that was not designed specifically to collect measures of teaching supports. Although the survey instruments I relied on were right enough to include several items asking about teaching supports, I still had to rely on data-driven methods to develop and validate the measures of teaching supports, leading me to rely on inference to determine what the items actually signal and which measures they align to. Future work can use the types of teaching supports that I use in this chapter and develop instruments that measure them directly, both with fewer and better items.

Finally, my cost-benefit analysis narrowly focuses on the financial costs and benefits associated with increasing teaching retention through teaching supports. Teacher retention, however, impacts school organizations beyond direct financial costs and benefits. For example, Ingersoll and Smith (2004) suggested that well-managed school organizations lead to reduced teacher attrition by eliminating low performers and bringing in new teachers who help facilitate instructional innovation. Conversely, high teacher attrition rates worsen the situation at school organizations that might be already struggling by undermining the "organizational stability, coherence, and morale" (p. 32) of the remaining teachers. These are all examples of non-financial costs of teacher turnover to a school. However, these non-financial costs are more difficult to quantify and include in this cost-benefit analysis. For this reason, future work should focus on expanding upon this current work to organizational outcomes beyond direct financial replacement costs for early career teachers and include more rich measures of its impacts on school organizations more generally.

# CHAPTER 4

# Development and Initial Validation of Measures of Teaching Supports from Statewide Surveys of Teachers

The first few years as a teacher are the defining period of their professional career. During this time, new teachers learn to carry instructional activities alongside managing a classroom, lesson planning, and working with others in their professional environment. Principals, instructional coaches, and colleagues usually mentor new teachers during this period, either through formal teacher induction activities or informally as part of their day-to-day interactions. These formal and informal activities have the potential to help beginner teachers transition into their new roles and start growing professionally. Yet, up to a third of new teachers leave teaching within their first five years (Ingersoll & Merrill, 2012). This trend is even more concerning once we consider that beginning teachers make a greater share of the workforce than ever before (Ingersoll & Merrill, 2012), and that the recent COVID-19 pandemic has exacerbated the growing shortage of new teachers entering the profession (Sutcher et al., 2019). As a result, we can likely expect the demand for teaching supports for new teachers to increase in the coming years, especially while the share of veteran teachers who can provide these supports for new teachers will likely decrease. Thus, finding out how we can best support these new teachers is becoming increasingly important.

Teacher induction programs have been one of the preferred policy interventions for providing professional development to early career teachers. For example, since the late 1990s and in response to a state-wide shortage of new teachers, California has required all new teachers to complete an induction program in order to receive a permanent teaching credential (Howe, 2006). In this context, teacher induction has shown promising results for the retention and quality of new teachers (Darling-Hammond, 1997).

Multiple reviews of the literature (Howe, 2006; Ingersoll et al., 2014; Ingersoll & Strong, 2011) have found positive associations between induction programs and teacher outcomes. Mentoring of early career teachers, in particular, has been found to be positively associated with teacher instructional practices (Davis & Higdon, 2008; Evertson & Smithey, 2000; Nevins Stanulis & Floden, 2009) and student achievement (Fletcher & Strong, 2009; Fletcher et al., 2008; Rockoff, 2008; Thompson et al., 2005). These studies are correlational, so we cannot infer causal connections between participating in an induction program and later outcomes; various forms of selection could explain these correlations. For example, it could be possible that more promising teachers choose to start their careers at schools where they know they will receive more or better teaching supports. This would lead to conflate participation in induction activities with other unobserved teacher decisions. Randomization of teachers to receive teaching supports could be a way to address these concerns.

To this point, the only randomized control trial that assigned early career elementary school teachers to a comprehensive induction program—including a formal mentoring program and multiple feedback cycles—found no effects on student outcomes nor employment decisions, even after treatment group participants reported taking part in more frequent induction activities than the control group participants (Glazerman et al., 2010; Isenberg et al., 2009). As part of a fidelity implementation analysis, the authors argued that there were substantial differences in induction supports between conditions, yet they acknowledge some variation across sites: "different programs emphasiz[ed] different approaches [to teacher induction...] such as orientation, assessment, professional development workshops, mentoring/peer coaching, small-group activities, and classroom observation" (Glazerman et al., 2010, p. 5). Though these experimental results seem to contradict the correlational evidence, Kang and Berliner (2012) raised some concerns about these null results of this work, suggesting that there may not have been enough contrast in the activities in which treatment and control teachers participated to lead to significant differences in outcomes. That is, the teachers in the control group may have also been exposed to substantial induction supports.

More recent work on teacher induction has teased out how different aspects of induction programs could support the development of early career teachers' practice. Ronfeldt and McQueen (2017) used seven questions from the School and Staffing Survey (SASS) to create a composite measure of the extensiveness of the supports available to SASS respondents. They found that induction intensity was associated with a reduction in both teacher attrition and mobility among early career teachers. When they examined specific types of supports, they found that early career teachers who participated in mentoring programs, in beginners' seminars, or experienced supportive communication supervisors were less likely to leave teaching than their colleagues were. While SASS includes a nationally representative sample of teachers in the United States, the seven indicator questions that Ronfeldt and McQueen (2017) used only asked participants to indicate whether they received a support or not. This allowed the authors to develop blunt measures of teaching supports available to early career teachers, but they acknowledged that more work is needed to develop more comprehensive measures, including ones that focus on the quality of supports that early career teachers experience.

Along the same lines, Smith et al. (2018) followed 62 early career middle school mathematics teachers during their first three years of teaching experience. They collected multidimensional measures of content-focused supports that included classroom observations, surveys, and interviews and developed measures for the dimensions of induction supports that these teachers experienced, including mentoring, professional development, administrator support, and professional community. They found that, while participants' teaching practice developed substantially during their first three years of teaching, this development was not associated with the kinds and levels of induction supports that they reported receiving during this same period. This paper shows the importance of multidimensional measures for teaching supports, along dimensions that are similar to the ones that I develop in this work. However, it is difficult to know whether their null results are because of their small sample size or because the teaching supports are actually independent of their outcome measures.

Finally, Auletto (2021) conducted a mixed-method study of induction practices for 700 teachers in Michigan. As part of the data collection, they surveyed teachers for different support practices that included indicators for collaboration with colleagues, mentoring and professional development. These different items were combined using principal component factor analysis (PCA) to develop aggregate measures for the support practices of interest (i.e., collaboration with colleagues, mentoring, and professional development). They found somewhat mixed results for the relationship between these measures and teacher outcomes (teacher satisfaction, career plans, and overall sense of support). For example, valuing a close relationship with a colleague was positively associated with overall satisfaction with teaching, but frequency of discussions with a colleague was not associated with the same outcome. Additionally, having experienced professional development that met early career teachers' needs was associated with overall satisfaction with teaching, but was not associated with plans to return to the same school district or to teach long term.

There are some possible methodological limitations with this study as well. In particular, the key assumption of PCA is that the construct being measured is unidimensional, and PCA will calculate the scores based on this assumption. However, it is possible that the constructs of interest are indeed multidimensional, leading to either mismeasurement of the construct of interest or loss of important alternative support dimensions. A more refined approach to measurement could help to assess the overall dimensionality of the constructs and to develop (multiple) measures that reflect these findings.

Considering all these previous studies together, it becomes clear that more work is needed to understand both the kinds of teaching supports that early career teachers experience during their first years of teaching and how these teaching supports vary across them. In this chapter, I develop and validate a framework for teaching supports available to early career teachers. In this framework, I focus on four types of teaching supports for early career teachers: opportunities for *professional learning & development*, *mentoring* cycles of classroom observation and feedback, *evaluative feedback* on teaching practice as part of the formal teacher evaluation process, and the general *working conditions* at each school site.

This chapter makes several contributions to our current understanding of early career teachers' experiences and the teaching supports that they receive. On the theoretical side, my work builds upon recent research (Redding et al., 2019; Smith et al., 2018) suggesting that teaching supports play an important role in early career teachers' instructional improvement and retention. For example, I measure the extent to which teachers found the professional learning activities in which they participated helpful for their development instead of observing whether they were offered professional development at their school. My teaching supports focus on the instructional systems that early career teachers experience has the potential to identify high leverage supports that can be used to develop further interventions or experiments. Moreover, focusing on early career teachers can help better identify the factors that contribute to the development of teaching practice by investigating the time of a teacher's career during which these factors could have the greatest impact. If we assume teaching supports influence teaching practice, we can also expect that early career teachers might benefit most from receiving these teaching supports, as their teaching practice is likely less developed and more malleable.

In addition, my work adds to our understanding of how teaching supports relate over time to teacher learning at different stages of the early career period. Most prior studies have relied on cross-sectional samples of teachers from either the SASS/TFS studies (e.g., Kang and Berliner, 2012; Ronfeldt and McQueen, 2017; Smith and Ingersoll, 2004) or self-collected samples from one specific states (e.g., Auletto, 2021; Wechsler et al., 2012). My work, instead, leverages a longitudinal survey program in Tennessee and develops measures for teaching supports over time for the same group of teachers. This opens up new methodological approaches (and potential challenges) to data analysis that are not possible with cross-sectional data.

On the practical side, there have been many calls to increase the supply of new teachers in response to projected teacher shortages, especially in STEM subjects (Ingersoll & Merrill, 2012; Sutcher et al., 2016a, 2016b). Increasing the number of new teachers entering the teaching workforce is only part of the solution to teacher shortages, as the retention of new teachers remains an ongoing problem (Guarino et al., 2006). Given the importance of retaining teachers, it is critical to identify the specific types of teaching support that predict higher retention rates. Developing better measures for induction supports in this chapter will allow me to also later test which ones predict retention and instructional effectiveness, thus offering important guidance for policy and practice reforms.

Finally, there is surprisingly little evidence on how teaching supports for early career teacher vary across teachers and schools. For example, Ronfeldt and McQueen (2017) is one of the few studies to report how the induction supports that SASS respondents reported receiving correlate with both their own personal characteristics and their school environments. In detail, they found that Black teachers, teachers working with more students identified as limited English proficient, or teachers working in smaller schools were more likely to report receiving more induction supports than their peers. While not directly reporting on variation of induction supports across teachers and schools, Wechsler et al. (2012) found that student achievement and teacher employment decisions are related to the school context that teachers experienced instead of the induction supports that they received. Similarly, Moore Johnson et al. (2012) and Hammerness and Matsko (2012) argued that school context that teachers experience matters for their satisfaction, leading to downstream effects on their students' achievement and teachers' employment decisions. Therefore, it is important to understand how and to what extent my measures of types of teaching supports vary between teachers and schools characteristics. These analyses can inform our understanding of to whom teaching supports are made available and which school settings are more supportive for early career teachers.

## Research Questions

The goals of this chapter are two-fold. First, I develop measures of teaching supports using a state-wide, longitudinal survey program. Second, I explore the variation in these measures in content areas, schools, and districts to describe the existing variation in these teaching supports that I can leverage in subsequent analyses.

Specifically, I ask these research questions:

RQ1. What measures for types of teaching supports can be developed using state-wide survey data?

RQ2. What are the psychometric properties of these measures of types of teaching supports?

RQ3. What is the variation in these types of teaching support measures in relation to teacher, school, and district characteristics?

## Methods

In this section, I describe methods for developing, estimating, and validating measures for latent constructs for teaching supports that I previewed in the introduction. Then, I describe the methods to explore the variation in these measures between individual teachers, schools, and districts.

## Research Context

Data for this chapter comes from the State Longitudinal Data System (SLDS) in Tennessee. These data bring together multiple data sources and give access to the universe of students and teachers present in Tennessee's public schools. I use multiple SLDS datasets to develop measures for the teaching supports available to early career teachers.

The Tennessee Educator Survey (TES) includes teacher survey responses to multiple yearly questionnaires. The Tennessee Department of Education has surveyed Tennessee teachers annually since 2012. The survey consists of a core set of questions that all respondents complete and a set of optional modules that are assigned depending on respondent's characteristics (e.g., teachers with fewer than three years of experience complete the early career module or teachers who moved schools complete the workforce module). The core module and the optional modules undergo annual revisions and additions, leading to questions changing year-to-year between survey administrations to address new data collection needs and revise unclear questions.

The Personnel Information Reporting System (PIRS) is an administrative dataset that includes information about teachers' years of experience, school assignment, and other personal characteristics. These data are the main source of student, teacher, and school characteristics.

## Sample

Table 4.1 reports the number of TES respondents for which I can calculate at least one factor score across academic years. In total, I calculate at least one factor score for 188,880 teacher-year observations, representing 32.4% of all teachers in my panel. When focusing on response patterns over time, the number of factor scores that I can calculate increases over time, starting from 14,839 responses (or 18.3%) in 2012 and ending at 35,754 responses (or 42.7%) in 2018. This pattern is because of the survey questions I selected and changes to their assignment to submodules in the early administrations of the survey and to the core teacher module in later administrations.

Table 4.2 reports the same counts by teachers' years of experience. The distribution of responses is more evenly distributed over teacher experience. On average, I can calculate at least one factor score for 33% of the teachers with two or more years of experience. Response rates for teachers with fewer than 2 years of experience are somewhat lower; I can calculate factor scores for 22.5% of first-year teachers and 29.3% of teachers with one year of experience.

Table 4.3 reports the characteristics of the survey respondents and compares this group with other teachers who have not responded. Overall, I find that survey respondents are significantly different from non-respondents both on their personal characteristics and in the schools where they work. Compared with non-respondents, survey respondents are 5.4 percentage points (p.p.) more likely to be women, 11 p.p. more likely to be White, about one year older, about one year more

Table 4.1: Survey Response Counts over Time

| Year | Respondents | Total Teachers | Percentage |
|---|---|---|---|
| 2012 | 14,839 | 81,070 | 18.3% |
| 2013 | 21,037 | 83,146 | 25.3% |
| 2014 | 22,536 | 82,862 | 27.2% |
| 2015 | 32,820 | 84,982 | 38.6% |
| 2016 | 28,169 | 84,471 | 33.3% |
| 2017 | 33,715 | 83,561 | 40.3% |
| 2018 | 35,764 | 83,682 | 42.7% |
| Total | 188,880 | 583,774 | 32.4% |

Table 4.2: Responses by Years of Experience

| Years of Experience | Respondents | Total Teachers | Percentage |
|---|---|---|---|
| 0 | 9,489 | 42,230 | 22.5% |
| 1 | 10,509 | 35,860 | 29.3% |
| 2 | 9,810 | 31,070 | 31.6% |
| 3 | 9,036 | 27,511 | 32.8% |
| 4 | 8,483 | 25,533 | 33.2% |
| 5 | 7,827 | 23,844 | 32.8% |
| 6 | 7,458 | 22,393 | 33.3% |
| 7 | 7,192 | 21,400 | 33.6% |
| 8 | 7,129 | 20,810 | 34.3% |
| 9 | 7,051 | 20,376 | 34.6% |
| 10 | 104,896 | 312,747 | 33.5% |
| Total | 188,880 | 583,774 | 32.4% |

Table 4.3: Differences Between Survey Respondents and Non-Respondents

| Variable | (1) All | (2) Resp. | (3) Non-Resp. | (4) Diff. | (5) E.S. |
|---|---|---|---|---|---|
| *Panel A: Teacher Characteristics* | | | | | |
| Women | 0.785 | 0.807 | 0.753 | 0.054 | 0.13*** |
| Asian | 0.006 | 0.005 | 0.007 | -0.002 | 0.023 |
| Black | 0.158 | 0.132 | 0.194 | -0.063 | 0.172 |
| Hispanic | 0.073 | 0.075 | 0.070 | 0.005 | 0.01** |
| Native American | 0.002 | 0.002 | 0.002 | 0.001 | 0.01** |
| Pacific Islander | 0.002 | 0.002 | 0.002 | 0.000 | 0.009 |
| White | 0.861 | 0.908 | 0.795 | 0.114 | 0.33*** |
| Other | 0.047 | 0.028 | 0.073 | -0.045 | 0.212 |
| Age | 44.404 | 44.823 | 43.780 | 1.043 | 0.08*** |
| Years of Experience | 6.905 | 7.549 | 6.019 | 1.530 | 0.41*** |
| Mathematics | 0.057 | 0.058 | 0.056 | 0.002 | 0.008 |
| Science | 0.055 | 0.053 | 0.058 | -0.004 | 0.019 |
| Special Education | 0.141 | 0.138 | 0.147 | -0.009 | 0.027 |
| Elementary | 0.716 | 0.763 | 0.652 | 0.111 | 0.24*** |
| English Language Arts | 0.264 | 0.250 | 0.284 | -0.034 | 0.076 |
| N | 121,150 | 70,118 | 51,032 | | |
| *Panel B: School Characteristics* | | | | | |
| Girls | 0.484 | 0.483 | 0.485 | -0.002 | 0.059 |
| Boys | 0.516 | 0.517 | 0.515 | 0.002 | 0.05*** |
| Asian | 0.020 | 0.019 | 0.021 | -0.001 | 0.050 |
| Black | 0.272 | 0.231 | 0.330 | -0.099 | 0.339 |
| Hispanic | 0.090 | 0.086 | 0.095 | -0.009 | 0.091 |
| Native American | 0.003 | 0.003 | 0.003 | 0.000 | 0.08*** |
| Pacific Islander | 0.002 | 0.002 | 0.002 | 0.000 | 0.03*** |
| White | 0.613 | 0.658 | 0.548 | 0.109 | 0.34*** |
| English Language Learners | 0.106 | 0.101 | 0.114 | -0.013 | 0.098 |
| Free or Reduced-Price Lunch | 0.489 | 0.487 | 0.493 | -0.006 | 0.026 |
| Immigrant | 0.022 | 0.019 | 0.027 | -0.008 | 0.187 |
| Migrant | 0.001 | 0.001 | 0.001 | 0.000 | 0.07*** |
| Special Education | 0.168 | 0.170 | 0.164 | 0.007 | 0.09*** |
| Gifted or Talented | 0.017 | 0.018 | 0.017 | 0.001 | 0.04*** |
| N | 118,320 | 70,079 | 48,241 | | |

Notes. School characteristics report the percentage for each variable.
Joint tests of significance (teacher characteristics $\chi^2(15) = 9881.628$, $p < 0.01$ and school characteristics $\chi^2(13) = 4228.289$, $p < 0.01$).
Robust standard errors in parentheses. * p < 0.05 ** p < 0.01 *** p < 0.001

experienced, and 11 p.p. more likely to be elementary school teachers. When focusing on the schools where they work, I find that survey respondents work in schools with more White students (by 11 p.p.), fewer Black students (by 9.9 p.p.), and fewer English Language Learners (by 1.3 p.p.). It is worth noting here that, while there are significant differences in other school characteristics, point estimates and effect sizes are small enough (i.e., fewer than 1 p.p.) for these differences to be considered negligible.

## Teaching Supports Measure Development

I developed the teaching support measures in three stages that involved both qualitative and qualitative methods. First, I selected survey items for their potential to measure teaching supports. The survey instruments were designed to collect data on a variety of topics, with later instruments including more than one thousand survey items over several modules. During this first part of the analysis, I read and selected items that could relate to teaching supports. Second, I used Exploratory Factor Analyses (EFA) to further select survey items and develop potential sub-factors for each teaching support. Third, I used Confirmatory Factor Analyses (CFA) to confirm the factor structure for each teaching support and estimate model fit indices for each sub-factor. I provide more detail for each step in the following sections.

### Survey Item Selection

The first stage in the development of latent factors involved identifying and selecting survey questions administered between 2012 and 2018 that could relate to experiences with teaching supports. These questions came from two, related surveys: the First to the Top (FTTT) survey administered between 2012 and 2014 and the Tennessee Educator Survey (TES) administered starting 2015. I read each survey and selected potential survey questions that asked for experiences with and views of teaching supports that teachers received during the school year. In this phase, I focused on questions that were either part of the teacher core module or submodules intended for teachers. For example, I excluded survey questions assigned to school administrators or other teaching support (e.g., school counselors). This identified roughly 600 questions across seven years of survey administrations for further processing.

I then matched questions over the different survey administration years. This process allowed me to match similar survey items while still accommodating year-to-year variation in item wording (e.g., addition or removal of some text). I started from the last survey administration year (i.e., 2018) and worked backwards. For each administration year, I used a machine learning method (i.e., natural language processing, NLP) to compare the item text content and to match it to the survey items that had the same content but not necessarily the same wording. I processed each survey item and calculated its textual distance from all survey items in the previous survey administration. I then classified as "matching" those survey items with textual distances above a given threshold—one that would allow for enough differences between two matched items while still minimizing mismatches of unrelated items. This process identified 108 overall survey items across survey administrations that I use in the later steps.

### Measure Development

I then developed latent factors measuring early career teachers' professional experiences using psychometric scale development methods (Worthington & Whittaker, 2006) starting from the set of survey items discussed above. All analyses were carried out with Mplus 8.7.

Psychometrics offers a set of statistical methods used to develop and validate measures for unobserved latent factors. First, I randomly assigned TES respondents to two sub-samples. I checked for balance on observed respondent characteristics (e.g., teacher preparation program, student, and school characteristics) to ensure that observed differences between the two sub-samples. These results suggested that there were no differences between the two samples in the observed variables. During this phase, I also estimated traditional measures of reliability and internal consistency (i.e., alpha and inter-item correlation) to assess whether my proposed items could form scales. All measures suggested adequate to good estimates for each subfactor.

After this preliminary step, I fit an exploratory factor analysis (EFA) for each survey administration year. EFAs allow for the exploration of how items are related to latent factors without making assumptions about the dimensionality of the measure (i.e., the number of sub-factors) or the relationships among the items. I used a weighted least square mean and variance adjusted (WLSMV) estimator with geomin rotation. The WLSMV estimator is appropriate when conducting an EFA with latent variables (Preacher & MacCallum, 2003) when declaring all survey items as categorical variables (Bollen, 1989).[1] I used an oblique rotation as it does not assume that sub-dimensions of teaching supports would be independent of each other. Said in another way, oblique rotations allow for sub-factors to be correlated with each other instead of imposing independence among the factors that the EFA extracts. I used a geomin[2] oblique rotation. This rotation provides a solution that minimizes the products of the variables' loadings (Haig, 2018) and this leads to a rotated solution like a solution from a confirmatory factor analysis. During this phase, I retained items that had loadings greater than 0.40 and without significant cross-loading on sub-factors (Worthington & Whittaker, 2006). I ran all these analyses on a random subsample of respondents (i.e., the EFA

[1]The assumption here is that there is a latent unobserved continuous variable that manifest itself in the categorical response variable. As an example, we could use a water bottle to try to measure the temperature of a room. The temperature is a continuous variable, but we would only observe whether the water in the water bottle is ice, liquid, or steam.

[2]Geomin is a specific case of the quartimin rotation that is only implemented in Mplus.

subsample). I handled missing data using pairwise correlations, which is typical for the WLSMV estimator (Muthen et al., 2015).

For each survey administration year, I estimated different EFAs with an increasing number of sub-factors, starting from 1 and adding as many factors as the number of items allowed (usually this was between 2 and 4 additional factors). I compared the various EFA solutions to balance model parsimony with the information extracted from the items. Intuitively, a one-factor solution is the most parsimonious solution, but it could disregard much of the variance in the items. On the other hand, a solution with as many factors as items captures the full variance in the items but does not offer any data reduction.

There are multiple ways to assess which EFA solution is the preferred factor structure. Traditionally, the Kaiser-Meyer-Olkin (KMO) test and the Bartlett's test of sphericity have been used to assess whether a set of items is suitable for factor analysis (P. Kline, 2014). However, these tests do not assess how well the EFA results fit the data, but only assess whether the items could be synthesized using a factor analysis. For this reason, I used the EFA scree plots alongside each model fit indices to decide which factor structure best represents the items. The scree plots are useful for visually inspecting the factor analysis eigenvalues and to find the point of scree. This point qualitatively identifies the ideal number of factors that balance the amount of variance extracted and the parsimony of factors. Model fit indices assess how well the proposed EFA model fits the item variance/covariance matrix (R. B. Kline, 2015). Current best practice is to report five fit indices to assess the goodness of fit of a model; each one provides slightly different information about model fit (Hu & Bentler, 1999; Marsh et al., 1988; Sivo et al., 2006). Fit indices are divided into two classes: absolute and incremental. Absolute fit indices assess the proposed model's departure from the perfect fit. A good fitting model has absolute fit indices below the accepted cut-off values (i.e., chi square statistics not significant, root mean square error of the approximation or RMESA $< 0.06$, and standardized root mean residual or SRMR $< 0.05$, Hu and Bentler, 1999). Incremental fit indices measure how well the proposed model reflects the best possible model. A good fitting model has incremental fit indices above accepted cut-off values (i.e., comparative fit index or CFI and Tucker-Lewis index or TLI $> 0.95$ Hu and Bentler, 1999). All CFAs used the WLSMV estimator for consistency with the EFA analyses.

### Validity Assessment

After I extracted the potential sub-dimensions for teaching supports, I used a series of confirmatory factor analyses (CFA) to test whether the factor structures extracted during the EFA are replicable in a new sample of FttT/TES respondents using the second random sample of respondents that I selected during phase one. The CFA framework provides a robust set of statistical tools to assess the internal validity and reliability of a measure. These analyses make the complementary assumptions to an EFA as they assume a factor structure and assess its fit to the available data.

I ran separate CFAs for each survey administration year and data availability group. This choice is a compromise between the ideal situation where all survey items are considered as part of the same factor analysis and the computational limitations with working with a large set of survey items and missing data patterns. Ideally, it is preferable to use a single CFA as this would estimate the full variance/covariance matrix among all survey items. In my case, this would result in a matrix of size 108 with roughly 12,000 covariance pairs, which makes this CFA unfeasible with the computational capacity I have access to. As a compromise, I have fit separate CFAs for each survey administration year and data availability group, leading to more manageable CFAs at the cost of making some assumptions about the covariance structure of the items. That is, fitting separate CFAs for each type of teaching supports is equivalent to assuming that there is no covariance with items that I categorized under different types of teaching supports.

Then, I calculated fit indices for all the CFAs. Recently, there have been some concerns with using traditional fit indices calculated using one sample of respondents (Maydeu-Olivares, 2017; Ximénez et al., 2022). I use a bootstrap procedure to address—at least partially—these concerns. This bootstrap procedure is preferable as fit indices estimate CFA fit to a specific sample, and they could be sensitive to sample selection bias. Bootstrapping (or resampling) can attenuate the risk of overstating the goodness of fit to a specific sample by estimating fit indices variation over multiple synthetic samples. In more detail, I used a bootstrap procedure to resample with replacement from the set of survey respondents. For each CFA, I drew 20 different samples and saved the fit indices for each draw. Finally, I report the distribution of these fit indices, using the median as the most likely value for each fit index.

**Overall Teaching Supports Measure.** Prior work has also treated teaching supports as a single construct, both theoretically (e.g., Glazerman et al., 2010; Isenberg et al., 2009) or methodologically (e.g., Ronfeldt, 2012). These single measures usually capture the intensity of a teacher induction program, like in the case of Glazerman et al. (2010)'s evaluation of a comprehensive teacher induction program, or as the overall number of induction supports that early career teachers received, like in Ronfeldt and McQueen (2017). To make my measures of teaching supports more comparable to this prior work, I developed a measure for the overall teaching supports that teachers report receiving. I developed this measure by combining the individual types of teaching support measure together in a single CFA for each survey administration year. Conceptually, this is similar to running a principal component factor analysis on the types of teaching supports I developed and taking the first component of them. I still report the same fit indices for this measure of the overall teaching supports as I described above for the individual types of teaching support measures.

# Results

In this section, I first review the results of development and validation of the teaching support measures. I then describe the variation in scores at the teacher, school, and district levels.

## Measure Development and Validity Assessment

### Exploratory Factor Analysis

The results of my initial EFAs eliminated some survey items that did not fit the estimated factor structure and identified four types of teaching supports suggested four types of teaching supports emerging from the survey items: professional learning & development, mentoring, evaluative feedback, and working conditions. The estimates of internal consistency for each of these types of teaching supports are presented in Table 4.4. Each column of the table reports the internal consistency estimates for one specific teaching support type, each row reports the estimates for the same survey administration year, Panel A reports estimates of the alpha coefficients for each construct and administration year, and Panel B reports the estimates of the inter-item correlation coefficient. Scanning the alpha coefficient estimates in Panel A, the minimum value is 0.741 and the maximum value is 0.965, with average values across all survey administration years above 0.800. These estimates suggest good internal consistency of the items that I selected for each teaching support type (Streiner, 2003). Another way to assess the internal consistency of a set of items is calculating the inter-item correlation coefficient (IIC). I report these estimates in Panel B. The minimum IIC coefficient is 0.240 and the maximum is 0.733, with all averages above 0.471. Although these estimates still show good internal consistency for each measure, they might also suggest redundancy of survey items for some scales (R. B. Kline, 2015).

### Confirmatory Factor Analysis

I then used confirmatory factor analysis (CFA) to assess the internal structure of each submeasure. Appendix Table A.3 reports factor loadings, sub-factors correlations, and item error correlations (when included in the analysis) for all CFAs. First, all factor loadings were significant, indicating that all survey items were related to the teaching support sub-factors I assumed they measured. Second, the average factor loading across all survey administrations is 0.741, suggesting good average loadings. Moreover, this average could be seen as a reliability estimate (Raykov, 1997), suggesting adequate average reliability for my measures.

Another way to assess model fit is through fit indices. The bootstrap fit indices for each CFA are reported in Table 4.5. I calculated these indices by estimating 20 CFAs, each on a bootstrapped sample from the original survey respondents. Each row in the table reports the mean fit index

Table 4.4: Internal Consistency Estimates

| | (1)<br>**PL&D** | (2)<br>**Eval. Feedback** | (3)<br>**Work Cond.** | (4)<br>**Mentoring** |
|---|---|---|---|---|
| *Panel A: Internal Consistency (Alpha Coefficient)* | | | | |
| Average | 0.874 | 0.847 | 0.800 | 0.914 |
| 2018 | 0.921 | 0.866 | 0.741 | 0.912 |
| 2017 | 0.902 | 0.814 | 0.784 | 0.916 |
| 2016 | 0.909 | 0.865 | 0.869 | 0.941 |
| 2015 | 0.871 | 0.899 | 0.808 | 0.889 |
| 2014 | 0.905 | 0.794 | | |
| 2013 | 0.851 | 0.866 | | |
| 2012 | 0.760 | 0.823 | | |
| *Panel B: Inter Item Correlation (IIC)* | | | | |
| Average | 0.471 | 0.587 | 0.501 | 0.670 |
| 2018 | 0.538 | 0.617 | 0.588 | 0.720 |
| 2017 | 0.605 | 0.421 | 0.475 | 0.733 |
| 2016 | 0.714 | 0.478 | 0.486 | 0.728 |
| 2015 | 0.458 | 0.641 | 0.457 | 0.499 |
| 2014 | 0.404 | 0.491 | | |
| 2013 | 0.342 | 0.763 | | |
| 2012 | 0.240 | 0.699 | | |

Table 4.5: Fit Indices

| **Index** | **2018** | **2017** | **2016** | **2015** | **2014** | **2013** | **2012** |
|---|---|---|---|---|---|---|---|
| *Panel A: Working Conditions and Mentoring* | | | | | | | |
| RMSEA | 0.034 | 0.029 | 0.046 | 0.034 | 0.148 | 0.126 | 0.115 |
| | [0.026, 0.042] | [0.025, 0.033] | [0.043, 0.049] | [0.032, 0.036] | [0.133, 0.163] | [0.116, 0.136] | [0.097, 0.133] |
| TLI | 0.998 | 0.998 | 0.99 | 0.991 | 0.989 | 0.993 | 0.994 |
| | [0.997, 0.999] | [0.997, 0.999] | [0.988, 0.992] | [0.990, 0.992] | [0.987, 0.991] | [0.992, 0.994] | [0.991, 0.997] |
| CFI | 0.999 | 0.999 | 0.992 | 0.993 | 0.991 | 0.995 | 0.996 |
| | [0.998, 1.000] | | [0.991, 0.993] | [0.992, 0.994] | [0.990, 0.992] | [0.994, 0.996] | [0.994, 0.998] |
| SRMR | 0.007 | 0.012 | 0.032 | 0.028 | 0.031 | 0.023 | 0.016 |
| | [0.004, 0.010] | [0.010, 0.014] | [0.029, 0.035] | [0.025, 0.031] | [0.027, 0.035] | [0.021, 0.025] | [0.014, 0.018] |
| *Panel B: Professional Learning & Development and Evaluative Feedback* | | | | | | | |
| RMSEA | 0.07 | 0.079 | 0.051 | 0.056 | 0.097 | 0.097 | 0.073 |
| | [0.068, 0.072] | [0.076, 0.082] | [0.047, 0.055] | [0.053, 0.059] | [0.094, 0.100] | [0.094, 0.100] | [0.069, 0.077] |
| TLI | 0.979 | 0.978 | 0.983 | 0.969 | 0.959 | 0.957 | 0.956 |
| | [0.978, 0.980] | [0.976, 0.980] | [0.980, 0.986] | [0.966, 0.972] | [0.956, 0.962] | [0.954, 0.960] | [0.951, 0.961] |
| CFI | 0.984 | 0.983 | 0.987 | 0.975 | 0.965 | 0.965 | 0.965 |
| | [0.983, 0.985] | [0.981, 0.985] | [0.985, 0.989] | [0.973, 0.977] | [0.963, 0.967] | [0.963, 0.967] | [0.961, 0.969] |
| SRMR | 0.054 | 0.044 | 0.056 | 0.065 | 0.057 | 0.047 | 0.039 |
| | [0.051, 0.057] | [0.042, 0.046] | [0.050, 0.062] | [0.061, 0.069] | [0.055, 0.059] | [0.045, 0.049] | [0.037, 0.041] |
| *Panel C: Overall Teaching Supports* | | | | | | | |
| RMSEA | 0.266 | 0.346 | 0.461 | 0.712 | 0.102 | 0.097 | 0.14 |
| | [0.262, 0.270] | [0.340, 0.352] | [0.450, 0.472] | [0.700, 0.724] | [0.094, 0.110] | [0.087, 0.107] | [0.125, 0.155] |
| TLI | 0.806 | 0.699 | 0.464 | 0.271 | 0.853 | 0.926 | 0.823 |
| | [0.802, 0.810] | [0.687, 0.711] | [0.446, 0.482] | [0.249, 0.293] | [0.829, 0.877] | [0.910, 0.942] | [0.790, 0.856] |
| CFI | 0.903 | 0.85 | 0.732 | 0.757 | 0.951 | 0.975 | 0.941 |
| | [0.901, 0.905] | [0.844, 0.856] | [0.723, 0.741] | [0.750, 0.764] | [0.943, 0.959] | [0.970, 0.980] | [0.930, 0.952] |
| SRMR | 0.085 | 0.082 | 0.132 | 0.132 | 0.066 | 0.074 | 0.104 |
| | [0.084, 0.086] | [0.079, 0.085] | [0.127, 0.137] | [0.129, 0.135] | [0.061, 0.071] | [0.066, 0.082] | [0.094, 0.114] |

Notes. RMSEA=Root Mean Square Error of Approximation; TLI=Tucker-Lewis Index; CFI=Comparative Fit Index; SRMR=Standardised Root Mean Residual.

from all bootstrapped repetitions and the minimum and maximum values in brackets. Four separate estimates of fit are reported: the mean square error of the approximation root (RMSEA), the Tucker-Lewis index (TLI), the comparative fit index (CFI), and the standardized root mean residual (SRMR). Customarily, it is preferred to have RMSEA values less than 0.05, TLI/CFI values greater than 0.95, and SMRM values less than 0.05 (Fan & Sivo, 2007; Hu & Bentler, 1999). Each column in the table reports the fit indices for separate survey administration years. Each panel of the table reports the fit indices for each teaching support (Panels A through C) and a factor analysis where all teaching supports are combined in a single measure of overall teaching supports (Panel D).

Scanning the results, all TLI/CFI values are above the customary cutoff value suggesting that the factor structure identified in the first stage of these analyses (i.e., the EFA) have identified an appropriate factor structure for the survey response patterns I observe.[3] Conversely, the RMSEA and SRMR estimates suggest that later survey administration years are a better fit to the data than earlier ones. Survey administrations from the Tennessee Educator Survey (TES) appear to be consistently a good fit to the data while earlier administrations of the First to the Top (FttT) survey might have less than ideal fit, suggesting possible heterogeneity in the survey responses either at the item level or across respondents' sub-groups.[4] This intuition also suggests that the year-to-year updates to the survey items could have helped with refining and refusing the survey items over time, leading to earlier survey items to have less than ideal psychometric properties.

Moving on to the fit indices for the combined measure of teaching support presented in Panel D, the fit indices suggest that a single combined measure of overall teaching supports might not be a good fit to the data, as all fit indices across all survey administration years do not meet the customary cutoff values for even decent fit of this overall measure to the data. In other words, while this measure of overall teaching supports extracts the principal component of the various types of teaching supports, the CFA results suggest that this overall measure does not fully capture the multidimensional nature of the teaching supports factors. The individual measure types still have independent variation that is not accounted for in the overall measure.

Appendix Table A.3 reports the final factor structure, factor loadings, factor covariances, and error term covariances.

## Measure Heterogeneity over Teacher, School, and District Characteristics

In this section I explore the potential heterogeneity in teaching supports by teachers, schools, and districts. Table 4.6 reports the results of the variance decomposition for each teaching support measure in each column along the four nesting levels in each row: local education agencies (LEAs),

[3]TLI and CFI are measures of global fit assessing the fit of the overall factor structure to the data.

[4]RMSEA and SRMR are measures of local fit assessing how well each item measures the proposed latent factor.

schools, teachers, and yearly scores. I report two related coefficients for each level. The first is the amount of variance explained in each measure at each level; the second is the percent of variance explained at each level in brackets.

### Variance Decomposition

In general and across most factors, the variation in teaching support scores I observe among teachers is about ten times the variation that I observe among schools and forty times the variation I observe among LEAs. Moreover, about half of the overall variation in scores is explained by the residual term, which captures changes from year to year in teaching supports or other unobserved contributions to these scores beyond the nesting of observations within teachers, schools, and LEAs.

Three patterns emerge when comparing differences in variance explained in each teaching support measure. First, the measure for mentoring has the greatest amount of year-to-year variation among the factors measured, suggesting that scores on this measure are less stable within the same teacher than others. Second, the measure for professional learning & development has the highest between-teachers variation, suggesting that this measure might be more stable than the others within the same teacher. Third, variance patterns in the overall teaching supports factor appear to mostly follow variance patterns in the other four factors, suggesting that this aggregate measure of teaching support might still be informative despite its poor psychometric properties.

**Teacher-Level Covariates.** Table 4.7 reports the heterogeneity of the measures across teacher characteristics. Moving across the table rows, I find heterogeneity in the average teaching supports that pre-tenure teachers (i.e., teachers in their first two years of teaching) and tenure eligible (i.e., teachers in their third to fifth years of teaching) report when compared to more experienced teachers. Consistent with what we might expect, I find that, on average, pre-tenure teachers report receiving higher levels of all four kinds of supports, with professional learning & development and evaluative feedback having the highest scores and mentoring and working conditions the lowest. At the same time, tenure eligible teachers report receiving less support across most teaching supports types, from roughly a third to a half of the amount reported by early career teachers. I observe the greatest difference in teaching supports between pre-tenure and tenure eligible teachers for working conditions, with this measure decreasing in magnitude by about two-thirds when comparing the average reported by pre-tenure and tenure eligible teachers.

Table 4.6: Variance Decomposition

| | (1) Prof. Learn. & Devel. | (2) Mentoring | (3) Eval. Feed. | (4) Work Cond. | (5) Overall Instr. Supp. |
|---|---|---|---|---|---|
| LEA intercept | 0.013 | 0.004 | 0.009 | 0.017 | 0.013 |
| | [1.6%] | [1.1%] | [1.2%] | [2.1%] | [1.4%] |
| School Intercept | 0.029 | 0.014 | 0.028 | 0.052 | 0.038 |
| | [3.6%] | [3.6%] | [3.5%] | [6.5%] | [4.1%] |
| Teacher Intercept | 0.341 | 0.118 | 0.302 | 0.276 | 0.404 |
| | [41.6%] | [29.4%] | [38.9%] | [34.5%] | [43.5%] |
| Residuals | 0.437 | 0.264 | 0.438 | 0.455 | 0.475 |
| | [53.3%] | [65.9%] | [56.4%] | [56.9%] | [51.1%] |

Notes. These estimates are from multilevel models nesting teacher-year teaching support measures within teachers (level 1), within schools (level 2), and within local educational agencies (level 3). Each cell reports the variance decomposition at each nested level. Percentage of total variance explained at each level is reported in brackets. LEA=local educational agency.

Table 4.7: Teaching Supports Heterogeneity along Teacher Characteristics

| | (1) PL&D | (2) Mentoring | (3) Eval. Feed. | (4) Work Cond. | (5) Overall |
|---|---|---|---|---|---|
| Pre-Tenure | 0.278*** | 0.134*** | 0.261*** | 0.119*** | 0.303*** |
| | (0.007) | (0.006) | (0.006) | (0.008) | (0.007) |
| Tenure Eligible | 0.086*** | 0.045*** | 0.095*** | 0.026** | 0.100*** |
| | (0.006) | (0.006) | (0.006) | (0.008) | (0.007) |
| Woman | -0.019** | -0.019** | 0.004 | -0.086*** | -0.011 |
| | (0.007) | (0.006) | (0.007) | (0.009) | (0.008) |
| Asian | -0.027 | 0.018 | -0.068 | 0.039 | -0.050 |
| | (0.047) | (0.043) | (0.046) | (0.059) | (0.050) |
| Black | 0.198*** | 0.054*** | 0.121*** | 0.036** | 0.163*** |
| | (0.010) | (0.009) | (0.010) | (0.013) | (0.010) |
| Hispanic/Latino | 0.067** | 0.010 | 0.051* | -0.081** | 0.061* |
| | (0.026) | (0.021) | (0.026) | (0.027) | (0.027) |
| Native American | -0.089 | -0.139* | -0.130 | -0.343*** | -0.157 |
| | (0.082) | (0.068) | (0.081) | (0.094) | (0.088) |
| Pacific Islander | 0.147 | 0.106 | 0.201* | 0.096 | 0.198 |
| | (0.101) | (0.091) | (0.099) | (0.125) | (0.108) |
| Other | -0.013 | -0.008 | 0.011 | -0.001 | -0.034 |
| | (0.173) | (0.138) | (0.174) | (0.181) | (0.185) |
| Constant | 0.122** | 0.085** | 0.047 | 0.227*** | 0.104* |
| | (0.038) | (0.032) | (0.037) | (0.045) | (0.042) |
| Endorsement Controls | Yes | Yes | Yes | Yes | Yes |
| School Controls | Yes | Yes | Yes | Yes | Yes |
| Peer Controls | Yes | Yes | Yes | Yes | Yes |
| Random Effects | Yes | Yes | Yes | Yes | Yes |
| $N$ | 167,307 | 103,237 | 167,307 | 103,237 | 168,594 |

Notes. All regressions in this table are 4-level multilevel models nesting teacher-year teaching support measures in teachers, schools, and school districts. School controls include the log number of students, standardized percentages of student demographic characteristics, standardized percentage of students identified as English Language Learners, students eligible for free or reduced priced lunch, students eligible for special education services or gifted and talented education. Peer controls include standardized measures of peers who are women, peers who identify with a specific racial/ethnic identity group, average peer years of teaching experiences, and peer teaching supports average.

Standard errors in parentheses. $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

Table 4.8: Teaching Supports Heterogeneity along Teacher Endorsement Areas

| | (1) PL&D | (2) Mentoring | (3) Eval. Feed. | (4) Work Cond. | (5) Overall |
|---|---|---|---|---|---|
| Elementary | 0.017 | 0.000 | 0.021* | 0.009 | 0.019* |
| | (0.009) | (0.007) | (0.008) | (0.010) | (0.009) |
| Math | -0.109*** | -0.059*** | -0.073*** | -0.085*** | -0.097*** |
| | (0.014) | (0.012) | (0.014) | (0.017) | (0.015) |
| Science | -0.116*** | -0.040*** | -0.119*** | -0.087*** | -0.136*** |
| | (0.014) | (0.012) | (0.013) | (0.017) | (0.015) |
| ELA | -0.023** | -0.018* | -0.016 | -0.022* | -0.020* |
| | (0.009) | (0.008) | (0.008) | (0.011) | (0.009) |
| Sp. Ed. | 0.007 | -0.003 | -0.004 | 0.002 | 0.005 |
| | (0.010) | (0.009) | (0.010) | (0.013) | (0.011) |
| Multiple Cred. | 0.015 | 0.023* | 0.007 | 0.034* | 0.010 |
| | (0.011) | (0.010) | (0.011) | (0.014) | (0.012) |
| Constant | 0.122** | 0.085** | 0.047 | 0.227*** | 0.104* |
| | (0.038) | (0.032) | (0.037) | (0.045) | (0.042) |
| Teacher Characteristics | Yes | Yes | Yes | Yes | Yes |
| School Controls | Yes | Yes | Yes | Yes | Yes |
| Peer Controls | Yes | Yes | Yes | Yes | Yes |
| Random Effects | Yes | Yes | Yes | Yes | Yes |
| $N$ | 167,307 | 103,237 | 167,307 | 103,237 | 168,594 |

Notes. All regressions in this table are 4-level multilevel models nesting teacher-year teaching support measures in teachers, schools, and school districts. School controls include the log number of students, standardized percentages of student demographic characteristics, standardized percentage of students identified as English Language Learners, students eligible for free or reduced priced lunch, students eligible for special education services or gifted and talented education. Peer controls include standardized measures of peers who are women, peers who identify with a specific racial/ethnic identity group, average peer years of teaching experiences, and peer teaching supports average.

Standard errors in parentheses. $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

### Teaching Supports Heterogeneity

**Teacher Characteristics.** Table 4.7 reports the heterogeneity of the teaching support measures over teacher characteristics. I find mixed results for differences between men and women in the teaching supports they report receiving. Only the mentoring coefficient is negative and significant, suggesting that women receive less mentoring. All others are not statistically significant, suggesting no differences between genders. Moving to race/ethnicity, Black teachers report receiving more support than their peers across all my teaching support measures. However, the coefficients for Hispanic/Latino teachers and Native American teachers are mostly negative and significant, suggesting that these groups of teachers report fewer teaching supports, except for teaching supports related to evaluative feedback.

Table 4.8 reports the heterogeneity of teaching supports across teacher endorsement areas. I find that teachers who are endorsed in mathematics, science, or English language arts report receiving lower levels of teaching supports than their colleagues. Teachers endorsed in other subjects/grades (i.e., elementary education or special education) or that hold multiple credentials report receiving similar teaching supports as the average teacher in the state.

**School-Level Student Characteristics.** Table 4.9 reports the heterogeneity of the instructional measures across student characteristics. Looking at school-level average student characteristics, I find that teachers who work at larger schools report lower level of support than their peers working at smaller schools, with an increase in one percent in the number of students corresponding to a decrease of about 0.03 standard deviations in teaching support measures. Teachers working in schools serving more Black or Hispanic/Latino, or students eligible for free or reduced price lunch report lower level of support than their peers. On the other hand, teachers who work with more gifted or talented students report more support than their peers do. Working in schools with more students eligible for other special education services is unrelated to the levels of supports that teachers report receiving.

**Peer Teacher Characteristics.** Previous work found that the settings in which teachers learn to teach make a difference in their later outcomes (Ronfeldt, 2012; Ronfeldt et al., 2015). It is possible that different settings might influence early career teachers' development differently. For this reason, I calculate average of peer characteristics and of teaching supports scores and include them in the regression. Table 4.10 reports the results of these analyses. Focusing on peer characteristics, I find that teachers who work in schools with more women and more experienced teachers report receiving higher levels of teaching supports. In contrast, teachers who work with more teachers who identify as Black, Hispanic, Latino, or other minoritized racial/ethnic identity groups report receiving less teaching supports.

Table 4.9: Teaching Supports Heterogeneity along School Characteristics

| | **(1) PL&D** | **(2) Mentoring** | **(3) Eval. Feed.** | **(4) Work Cond.** | **(5) Overall** |
|---|---|---|---|---|---|
| log(N students) | -0.025$^{***}$ | -0.016$^{***}$ | -0.018$^{**}$ | -0.031$^{***}$ | -0.024$^{***}$ |
| | (0.006) | (0.005) | (0.005) | (0.007) | (0.006) |
| Asian % (std) | -0.005 | 0.002 | -0.005 | 0.005 | -0.007 |
| | (0.004) | (0.003) | (0.004) | (0.004) | (0.004) |
| Black % (std) | 0.011 | 0.003 | -0.001 | -0.012 | 0.002 |
| | (0.008) | (0.006) | (0.007) | (0.009) | (0.008) |
| Hisp./Lat. % (std) | -0.004 | -0.006 | -0.007 | -0.016$^{**}$ | -0.007 |
| | (0.005) | (0.004) | (0.005) | (0.006) | (0.006) |
| Native Am. % (std) | 0.003 | 0.004 | 0.000 | 0.004 | 0.001 |
| | (0.003) | (0.003) | (0.003) | (0.004) | (0.003) |
| Pacific Isl. % (std) | 0.006 | 0.000 | 0.004 | 0.004 | 0.007 |
| | (0.004) | (0.004) | (0.004) | (0.006) | (0.004) |
| ELL % (std) | -0.003 | -0.006 | -0.001 | -0.007 | -0.001 |
| | (0.005) | (0.004) | (0.005) | (0.006) | (0.005) |
| FRPL % (std) | -0.008$^{*}$ | -0.011$^{**}$ | -0.012$^{**}$ | -0.017$^{***}$ | -0.014$^{***}$ |
| | (0.004) | (0.003) | (0.004) | (0.005) | (0.004) |
| Sp. Ed. % (std) | 0.006 | 0.002 | 0.006 | -0.000 | 0.007 |
| | (0.004) | (0.004) | (0.004) | (0.005) | (0.004) |
| Gifted % (std) | 0.007$^{*}$ | 0.005 | 0.007$^{*}$ | 0.009$^{*}$ | 0.009$^{*}$ |
| | (0.003) | (0.003) | (0.003) | (0.004) | (0.004) |
| Constant | 0.122$^{**}$ | 0.085$^{**}$ | 0.047 | 0.227$^{***}$ | 0.104$^{*}$ |
| | (0.038) | (0.032) | (0.037) | (0.045) | (0.042) |
| Teacher Controls | Yes | Yes | Yes | Yes | Yes |
| Peer Controls | Yes | Yes | Yes | Yes | Yes |
| Random Effects | Yes | Yes | Yes | Yes | Yes |
| $N$ | 167,307 | 103,237 | 167,307 | 103,237 | 168,594 |

Notes. All regressions in this table are 4-level multilevel models nesting teacher-year teaching support measures in teachers, schools, and school districts. Teacher controls include indicators for career stage, demographic characteristics, and subject endorsements. Peer controls include standardized measures of peers who are women, peers who identify with a specific racial/ethnic identity group, average peer years of teaching experiences, and peer teaching supports average.

Standard errors in parentheses. $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

Focusing on peer teaching support measures, I find that these peer teaching support measures are weakly predictive of teachers' own scores for all teaching support measures. This finding could suggest that individual experiences with teaching supports might still be independent of what others at the same school experience. However, the point estimates suggest that the working conditions experienced by individual teachers are strongly related to the average working conditions that their peers reported. The same relationships for the other three types of teaching supports have similar in sign but smaller in magnitude point estimates, suggesting that working conditions that individual teachers experience are more strongly related to their peers', while other types of teaching supports are more individualized to each teacher.

# Discussion

Every teacher was a new teacher at some point of their career. For this reason, there has long been an interest in understanding how to best support the development of teaching practice during the early career. Prior work on this topic has shown that most new teachers participate in some sort of induction program. However, it is not clear how these induction programs help with the development of practice or about the overall variation in these teaching supports throughout the teacher population. In this chapter, I developed a new set of measures for teaching supports using teacher survey data from Tennessee, and then I found variation in these measures across teacher and school characteristics.

## Teaching Supports Measures

The factor structure I extract and validate in this chapter highlights several properties of my measures of teaching supports. First, I found that one general factor for overall teaching supports is a worse fit to the data than separate latent factors for each type of teaching supports. This suggests that teachers report enough variation in each survey across the teaching supports types to warrant different factors rather than a single combined measure for overall support.

This result is in line with recent work that has tried to develop multidimensional measures for teaching supports (Auletto, 2021; Smith et al., 2018) and a break from prior work that relied on combined measures of support intensity (e.g., Ronfeldt and McQueen, 2017). Moreover, these results suggest that multidimensional measures of teaching supports can be developed from existing data collection efforts if the survey instruments are rich enough to measure variation along multiple dimensions. Both the FttT and TES instruments included enough survey items to identify questions that measure multiple dimensions of teaching supports. This is in contrast to, for example, the SASS survey instruments that only included less than ten questions about induction

Table 4.10: Teaching Supports Heterogeneity along Peer Characteristics

| | (1) PL&D | (2) Mentoring | (3) Eval. Feed. | (4) Work Cond. | (5) Overall |
|---|---|---|---|---|---|
| Peer Women (std) | 0.012** | 0.012*** | 0.017*** | 0.010* | 0.018*** |
| | (0.004) | (0.003) | (0.004) | (0.004) | (0.004) |
| Peer Asian (std) | 0.008 | -0.002 | 0.003 | -0.001 | 0.004 |
| | (0.005) | (0.005) | (0.005) | (0.007) | (0.006) |
| Peer Black (std) | -0.018* | -0.021** | -0.017* | -0.028** | -0.019* |
| | (0.008) | (0.007) | (0.007) | (0.010) | (0.008) |
| Peer Hisp./Lat. (std) | -0.006* | -0.007** | -0.010** | 0.001 | -0.012*** |
| | (0.003) | (0.003) | (0.003) | (0.003) | (0.003) |
| Peer Native (std) | 0.001 | 0.002 | 0.002 | 0.005 | 0.001 |
| | (0.003) | (0.002) | (0.003) | (0.003) | (0.003) |
| Peer Pacific Isl. (std) | -0.005 | -0.002 | -0.006* | -0.001 | -0.006* |
| | (0.003) | (0.002) | (0.003) | (0.003) | (0.003) |
| Peer Other (std) | -0.013* | 0.001 | -0.014** | -0.006 | -0.025*** |
| | (0.005) | (0.005) | (0.005) | (0.006) | (0.006) |
| Peer Yrs. Exp. (std) | 0.026*** | 0.023*** | 0.023*** | 0.035*** | 0.035*** |
| | (0.005) | (0.005) | (0.005) | (0.006) | (0.005) |
| Peer PL&d | 0.133*** | | | | |
| | (0.002) | | | | |
| Peer Mentoring | | 0.110*** | | | |
| | | (0.002) | | | |
| Peer Eval. Feed. | | | 0.129*** | | |
| | | | (0.002) | | |
| Peer Work Cond. | | | | 0.222*** | |
| | | | | (0.003) | |
| Peer Overall | | | | | 0.148*** |
| | | | | | (0.002) |
| Constant | 0.122** | 0.085** | 0.047 | 0.227*** | 0.104* |
| | (0.038) | (0.032) | (0.037) | (0.045) | (0.042) |
| Teacher Controls | Yes | Yes | Yes | Yes | Yes |
| School Controls | Yes | Yes | Yes | Yes | Yes |
| Random Effects | Yes | Yes | Yes | Yes | Yes |
| $N$ | 167,307 | 103,237 | 167,307 | 103,237 | 168,594 |

Notes. All regressions in this table are 4-level multilevel models nesting teacher-year teaching support measures in teachers, schools, and school districts. Teacher controls include indicators for career stage, demographic characteristics, and subject endorsements. School controls include the log number of students, standardized percentages of student demographic characteristics, standardized percentage of students identified as English Language Learners, students eligible for free or reduced priced lunch, students eligible for special education services or gifted and talented education.

Standard errors in parentheses. $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

supports (Kang & Berliner, 2012; Ronfeldt & McQueen, 2017; Smith & Ingersoll, 2004). Nevertheless, I still developed an overall teaching support measure that combines the types of teaching support measures into a single factor. While this overall measure only had modest psychometric properties, prior work has mostly relied on single measures of the intensity of induction supports (e.g., Ronfeldt and McQueen, 2017) or combined multiple survey items using principal component factor analysis (e.g., Auletto, 2021). In order to keep my results comparable to this prior work, I decided to develop and report on this overall measure.

Additionally, this work also shows some affordances and limitations in using secondary survey data for research purposes. Regarding affordances, I was able to identify multiple dimensions of teaching supports that align with existing literature on early teacher support. The supports I identified through factor analyses align theoretically with dimensions of teaching support that prior literature has deemed to be important, even though the supports come from survey instruments that were not specifically designed to measure these dimensions. These secondary data also provide me with a long panel of survey data. This is important as this allows me to observe teachers responding to multiple surveys, on average, three times. This helps with both addressing concerns with survey non-response and provides a data panel I can later use in other analyses.

At the same time, the choice to retrofit existing data to measure teaching supports comes with compromises. First, fit indices are lower than I wished for some years. This problem is particularly present for early survey administrations that used the First to the Top (FttT) survey instrument. These early administrations mostly focused on teachers' feedback on the new teacher evaluation process, leading to a narrow focus in the instructional measures that I can develop from the FttT survey administrations. That is, I am only able to calculate factor scores for the professional learning & development and evaluative feedback factors from the FttT instrument. Moreover, these factors still have less than adequate fit indices, suggesting low reliability of some items in the FttT survey instruments. In fact, some sections of the teacher survey measuring these factors were updated and rewritten during the switch to the Tennessee Educator Survey (TES) in 2015. While this revamping of the survey instrument improved the overall reliability of the instrument, as fit indices' improvement from TES suggests, this change also introduced a break in the panel for the professional learning & development and evaluative feedback factors.

A second limitation in using secondary teacher survey data is that the relationship between latent constructs and survey items is lower than what is commonly observed in survey instruments that were developed for specific research purposes. For example, some survey items (e.g., item "Indicators from the teacher observation rubric are often referenced in informal discussions between teachers" as a measure of mentoring) have lower than ideal factor loadings, suggesting low signal-to-noise ratios in these survey responses. As an ensemble, however, the items appear to sufficiently correlate with the extracted latent factors which allows me to retain them in the factor analysis regardless of their lower than ideal loadings.

Finally, there are dimensions of teaching supports that are missing from my analyses. For example, prior literature has suggested that school leaders and relationships with families play a crucial role in school improvement (Bryk et al., 2009; Moore Johnson, 2020). However, the EFA stage eliminated survey items that asked about school leadership or school climate because of poor fit of these questions to possible latent constructs. Future measurement work on teaching supports could address this limitation by developing and validating measures for these missing sub-dimensions of teaching supports.

## Teaching Supports Variation

I found that teacher-level variance (i.e., stability in scores within the same teacher and differences across teachers) explains about ten times the amount of variance in teaching support scores than differences across schools and about forty times the amount of variance across districts. To my knowledge, no prior work has reported variation along these dimensions. In more detail, year-to-year and within-teacher variation explain about an equal amount of variation and, together, they explain up to about 95 percent of the total variation in teaching support scores. First, this suggests that my teaching support measures capture teacher-level variation in the kinds of teaching support measures rather than school- or LEA-level differences in support that is shared by groups of teachers. In other words, there is more within-school rather than between-school variation in terms of average teaching supports offered to teachers, as I observe the majority of variation in teaching supports to be between teachers within the same schools. Second, about half of the variation in the teaching support measures remains in the residual term (i.e., unexplained) after accounting for teacher, school, and LEA variance components. This residual term captures within-level, year-to-year variation in teaching support measures (i.e., good variance), measurement error, other random error components (i.e., bad variance). This finding has methodological implications for analyses that consider teachers' evaluation and employment decisions as they justify using within-teacher analyses with these measures of instructional variance. On the policy side, this finding suggests that interventions aiming to improve school-level teaching supports should focus on providing adequate levels of support to all teachers within the same school, rather than focusing on variation in teaching supports across schools. This would also ensure equity among teachers in the kind and intensity of teaching supports that teachers experience.

Looking at differences across measures in variance decomposition, the measure for working conditions has about two times the variance explained at the school level than other factors that measure teacher-level supports. Intuitively, this might be expected, as the working conditions that one teacher reports receiving should be related to the working conditions that other teachers at the same school report. In other words, my results suggest that, of the four types of teaching supports I develop, working conditions is the one that relates the most to the experiences of other teachers at the same school site. This suggests that, at least at face value, my measures of teaching supports follow expected variation patterns. Moreover, individual and peer reports of working conditions are more strongly correlated than the other factors, highlighting again the importance of colleagues in shaping working conditions for individual teachers. Finally, these findings provide some validity evidence for the fact that the other three types of teaching supports (i.e., professional learning & development, mentoring, and evaluative feedback) might capture teacher-level support processes, while working conditions might also capture larger school-level support processes beyond what is offered to individual teachers.

The mentoring factor has the greatest unexplained variance among all my latent factors. This suggests that this kind of teaching support has the greatest year-to-year variation compared with other factors. Moreover, there is a lot of variation in the quality of mentoring that early career teachers experience. This observation is in line with prior work that has found that most new teachers are assigned a mentor (Ingersoll & Strong, 2011; Ronfeldt & McQueen, 2017) and that there is significant variation in the intensity and quality of mentoring that these new teachers report receiving (Matsko et al., 2007).

In focusing on predictors of teaching supports, I found that early career teachers report, on average, receiving more teaching support across most factors than their more experienced colleagues; early mid-career teachers report receiving higher support only in mentoring and professional development factors. Two possible mechanisms could explain these results. A more hopeful explanation is that school administrators and other teaching support personnel focus their support efforts on early career teachers and that this process is captured in my measures. This explanation is in line with the fact that most schools offer induction support for early career teachers during the first few years of teaching (Ingersoll & Strong, 2011).

A more cynical view of my results is that early career teachers perceive the same teaching supports as more helpful to their practice than their more experienced peers because, for example, it is the same time they receive that support, leading them to rate the same practices as more helpful as their colleagues. This explanation is in line with the intuition that data collection that relies only on self-reports could be biased if the respondents construe the survey items differently. Regrettably, I do not have a way to test if either of these explanations are true, as I do not have a way to observe teaching supports offered to teachers directly. Future work using my measures for teaching supports should try to account for individual-level differences in their reports (e.g., with teacher fixed effects) to address part of the concerns with self-reported data.

The variation in teaching support measures across teacher characteristics is more difficult to interpret. Black teachers report receiving higher support, while Hispanic/Latino and Native American teachers report lower support. This finding is similar to what Ronfeldt and McQueen (2017) found in the SASS data for Black teachers and opposite in sign for Hispanic/Latino teachers. However, it is difficult to know what mechanism is behind these differences. It is possible that Black teachers tend to work in schools where more induction supports are available, confounding individual teachers' characteristics with contextual school supports (Matsko et al., 2007). However, these differences are still present even after I account for school-level differences in student body composition and for a school-level random effect, which should in part adjust the estimates for the school-level average, suggesting that contextual school supports might not fully explain these differences.

Similarly, variation in student characteristics suggests that teachers who work in schools that serve a less diverse student body and a more economically advantaged one report receiving more teaching support than their colleagues working in different schools. These results are similar in sign to prior work (Loeb et al., 2005; Ronfeldt & McQueen, 2017), suggesting that teachers working in more advantaged contexts report experiencing higher levels of teaching supports. Again, I cannot separate teachers' perceptions of the supports that they receive from the quality of supports they actually received, so it is possible that teachers working in different schools might report feeling more or less supported even if they actually received the same teaching supports.

Overall, the results from the exploratory and confirmatory factor analyses and the multilevel models suggest that my measures of teaching supports have adequate psychometric properties to warrant their use in further analyses. Moreover, the analysis of variance suggests that I have enough year-to-year variation within teachers to warrant their use in a panel. Finally, the inclusion of teacher- and school-level covariates only explains teacher- and school-level variance, leading to year-to-year variation to become more salient.

# CHAPTER 5

# Implications for Practice, Policy, and Research

This dissertation has several implications for practice, policy, and research. Overall, I found that teaching supports matter for the instructional growth and employment decisions of early career teachers. Therefore, investment in teaching supports, especially at the state level, can help address the needs of early career teachers and increase retention. Finally, future research should focus on a causal evaluation of teaching supports, including identifying which teaching supports lead to an improvement in early career outcomes and an analysis of the mechanisms behind such improvements. I discuss each one of these points in more detail in the sections below.

## Teaching Supports Matter for Early Career Development

The results across the chapters in this dissertation suggest that early career teachers that report receiving more teaching supports develop at a faster rate and are less likely to turnover than their colleagues experiencing average teaching supports. In more detail, I find that teachers who report receiving enhanced teaching supports[1] develop about thirty-three percent faster than their peers on observation ratings, leading them to become eligible for tenure one year before their colleagues (three years vs. four years). The results for teacher value added to student test scores are less clear, but some models still suggest a small, positive improvement in the growth for early career teachers experiencing enhanced teaching supports equivalent to an additional eleven instructional days. Focusing on teacher retention, my results suggest that enhanced teaching supports increases the probability of retention in the same teaching position by twenty-five percent. Moreover, I find that instructional improvement as a result of experiencing enhanced teaching supports result in an improvement in teacher retention and that this pathway explains at least one fifth of the overall relationship between teaching supports and teacher retention.

Here it is also important to point out that my results compare teachers who reported enhanced

[1]Technically, these teachers report experiencing one standard deviation above the mean in teaching supports. I use *enhanced teaching supports* here to simplify the language in this chapter.

teaching supports to teachers who experience average teaching supports. This means that the counterfactual for my results is teachers still experiencing an average level of teaching supports. This means that the converse of my results are also possible. That is, teachers who report receiving diminished teaching supports[2] will develop at slower rates than their peers, putting them at risk of never becoming eligible for tenure, as most temporary teacher licenses expire after five years; this makes them more likely to leave their teaching position year after year. My results, therefore, point towards the need to strengthen program offering teaching supports and to develop new ones where not already present.

At the same time, this work opens up new questions and direction for future research into teaching supports. First, my measures of teaching supports improve upon prior work that only used blunt measures for them (e.g., Ronfeldt and McQueen, 2017) or measures from a single cohort of teachers (e.g., Smith et al., 2018) by developing more detailed, longitudinal measures of teaching supports. However, I am not able to describe how enhanced teaching supports are different from average teaching supports in terms of kinds of supports that teachers receive, the quantity and quality of the supports over the school year. This remains an open question that future research should address. For example, such work could describe what features of mentoring result in teachers reporting enhanced mentoring supports instead of average mentoring supports, whether the quantity of mentoring activities plays a role in enhanced mentoring supports, or if the quality of the mentoring activities makes a difference. Related to this point, between-teacher variation in teaching supports explains the greatest percentage of variance across all of my measures. One of the most likely explanations is that individual teachers report receiving consistent teaching supports and, perhaps, suggesting that their experiences with teaching supports depend on their teacher preparation experiences, personal characteristics, and current career goals. Future research should focus on describing how teacher characteristics affect teachers' experiences with teaching supports in more detail and on understanding how to design a teaching supports program that meets the needs of most teachers. For example, future work could interrogate whether teaching supports programs are more effective when they adapt to teachers' characteristics (e.g., completing a traditional teacher preparation program versus entering teaching through alternative teacher certification route) or whether a one-size-fits-all teaching supports program has similar effects on early career teachers' instructional effectiveness and employment decisions.

## Investment in Early Career Teaching Supports

My results show that teachers in their first three years of teaching benefit the most from enhanced teaching supports, suggesting that investment in teaching supports for this group of teach-

[2]That is, one standard deviation below the state average.

ers is likely to yield more benefits than investing on the same level of supports to all teachers. Therefore, a critical policy decision is who should develop and support investment in early career teachers. My work suggests that a large teaching supports program at the state level is likely to yield the biggest returns when compared to programs at the district or school level. A few considerations go into this conclusion. First, early career teachers are highly mobile; they are about twice as likely to move to a different school or school district than their more experienced colleagues, while remaining employed as a teacher within the same state. This means that individual schools or school districts might lose out on their investment in teaching supports when an early career teacher moves. Given that these teachers tend to remain in the same state labor market, however, investment in teaching supports at the state level would continue to reap benefits even after early career teachers move to different schools or districts. This suggests that state department of education should encourage and support teaching supports programs for early career teachers, as a way to defray some potential costs to schools or districts associated with providing enhanced levels of teaching supports.

The cost-benefit analysis that weighs the improved retention rates associated with participating in a teaching supports program against the costs associated with replacing teachers who are not retained suggested that, even for labor markets where the cost of replacement for a teacher is high, the overall budget for a teaching supports program should be around $2,000, which is well below the cost of high-quality teacher professional development programs. This result suggests that the best way to structure early career teaching supports programs might be through a cost-sharing setup between schools, school districts, and state departments of education. This insight is predicated on two findings from my work. While individual schools are more likely to benefit from better instruction from teachers who participate in the program, schools are also more likely to lose out on these gains if a teacher leaves for a different teaching position even within the same district. A cost-sharing setup between individual schools, districts, and state departments could help address this concern if the state provides bridge funding between the expected break even point for individual schools and the actual cost of an enhanced teacher supports programs. States should consider providing developing cost sharing setups where schools and school districts are expected to partially cover the costs of enhanced teaching supports and the state covers the rest. This setup would more equitably divide the costs associated with providing teaching supports to early career teachers among the stakeholders that are more likely to benefit from this investment.

This kind of cost-sharing setup, however, requires ongoing monitoring of the kinds of teaching supports that teachers receive. Though most of the variation in supports is between teachers in the same schools, there is still substantial variation in supports between schools and districts. Future research should describe what is the source of this variation between schools and districts and how approaches to the provision of teaching supports may vary school to school and district to district. Moreover, such work could be used to assess the implementation of teaching supports programs, given that there is variation between schools and districts in teachers' reports of teaching supports. Related, monitoring of teachers' experiences can also inform schools and districts of possible areas of improvement regarding their approaches to offering teaching supports. Finally, monitoring at the state level can ensure equitable teaching supports experiences across schools and school districts, guaranteeing that teachers who move between schools or districts continue to receive similar teaching supports. This may be especially important for schools that have larger enrollments and that serve a higher proportion of students eligible for free or reduced price lunch, where teachers tend to report lower levels of teaching supports.

## On the Nature of Teaching Supports

A final set of insights from this dissertation is about the nature of teaching supports, which can inform the program design of teaching supports programs aiming to improve teaching quality and retention. First, I find that a minimum of three years of sustained, enhanced teaching supports are necessary to observe appreciatively different teacher effectiveness outcomes. This suggests that current teacher induction programs—which average between one and two years of support—might be too short to result in meaningful improvements in teaching effectiveness. One area for future research, and possibly a causal analysis, could be the development and evaluation of a three-year teacher induction program versus a business-as-usual, one- or two-year teacher induction program.

Relatedly, designing and implementing multi-year programs to provide teaching supports will require more research in the needs that teachers have at different points of their careers. Year-to-year variation in measures of teaching supports suggests that the levels of teaching supports that early career teachers report receiving are greatest during the first two years of teaching but, by their third year of teaching, decline to average levels. These results signal that current teaching supports are focused on teachers in their first two years of experience when these supports are likely to have the greatest impact on instructional effectiveness growth and teacher retention. At the same time, there is a need to better understand what kinds of ongoing teaching supports may be necessary for teachers past their first and second year of experience to sustain instructional growth and retention.

# APPENDIX A

# Additional Tables

Table A.1: Heterogeneity by Teaching Support Type on Observation Ratings

| | (1)<br>**Overall Support** | (2)<br>**PL&D** | (3)<br>**Mentoring** | (4)<br>**Eval. Feedback** | (5)<br>**Work Cond.** |
|---|---|---|---|---|---|
| Factor | $0.049^{***}$ | $0.045^{***}$ | $0.040^{***}$ | $0.045^{***}$ | $0.032^{***}$ |
| | (0.004) | (0.005) | (0.008) | (0.005) | (0.006) |
| Factor × Pre-Tenure | $0.009^{***}$ | $0.010^{***}$ | 0.005 | $0.010^{***}$ | 0.002 |
| | (0.003) | (0.003) | (0.005) | (0.003) | (0.004) |
| Factor × Tenure Eligible | 0.001 | 0.001 | -0.005 | 0.001 | -0.002 |
| | (0.002) | (0.002) | (0.003) | (0.002) | (0.002) |
| Factor × Mid-Career | -0.001 | -0.001 | 0.001 | -0.001 | 0.001 |
| | (0.001) | (0.001) | (0.002) | (0.001) | (0.001) |
| Factor × Late Career | 0.000 | 0.000 | $0.001^{*}$ | 0.000 | $0.001^{**}$ |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Pre-Tenure | $0.145^{***}$ | $0.144^{***}$ | $0.135^{***}$ | $0.144^{***}$ | $0.138^{***}$ |
| | (0.004) | (0.004) | (0.005) | (0.004) | (0.005) |
| Tenure Eligible | $0.035^{***}$ | $0.035^{***}$ | $0.035^{***}$ | $0.035^{***}$ | $0.036^{***}$ |
| | (0.003) | (0.004) | (0.004) | (0.004) | (0.004) |
| Mid-Career | $0.008^{*}$ | $0.009^{**}$ | 0.003 | $0.009^{**}$ | 0.004 |
| | (0.003) | (0.003) | (0.004) | (0.003) | (0.004) |
| Late Career | $0.010^{***}$ | $-0.009^{**}$ | $-0.017^{***}$ | $-0.009^{**}$ | $-0.017^{***}$ |
| | (0.003) | (0.003) | (0.004) | (0.003) | (0.004) |
| Constant | $3.688^{***}$ | $3.678^{***}$ | $3.804^{***}$ | $3.678^{***}$ | $3.792^{***}$ |
| | (0.036) | (0.037) | (0.039) | (0.037) | (0.039) |
| Observations | 528,149 | 524,597 | 330,174 | 524,597 | 330,174 |
| F | 7.129 | 5.656 | 1.826 | 5.656 | 3.373 |
| *df* | 4.000 | 4.000 | 4.000 | 4.000 | 4.000 |
| p | 0.000 | 0.000 | 0.121 | 0.000 | 0.009 |

Notes. Experience is modeled using a linear spline with nodes at different career stages: pre-tenure teachers have between zero and two years of experience; tenure eligible teachers have between three and five years of experience; mid-career teachers have between six and ten years of experience; late career teachers have more than ten years of experience.

Robust standard errors in parentheses. $^{*}$ $p < 0.05$ $^{**}$ $p < 0.01$ $^{***}$ $p < 0.001$

Table A.2: Heterogeneity by Teaching Support Type on TVAAS

| | (1) Overall Support | (2) PL&D | (3) Mentoring | (4) Eval. Feedback | (5) Work Cond. |
|---|---|---|---|---|---|
| Factor | 0.000 | -0.008 | -0.071 | -0.008 | -0.040 |
| | (0.019) | (0.021) | (0.042) | (0.021) | (0.031) |
| Factor × Pre-Tenure | 0.000 | 0.005 | 0.027 | 0.005 | 0.012 |
| | (0.012) | (0.013) | (0.026) | (0.013) | (0.019) |
| Factor × Tenure Eligible | 0.001 | -0.002 | 0.012 | -0.002 | 0.013 |
| | (0.006) | (0.007) | (0.013) | (0.007) | (0.010) |
| Factor × Mid-Career | -0.002 | -0.000 | -0.005 | -0.000 | -0.003 |
| | (0.004) | (0.004) | (0.008) | (0.004) | (0.006) |
| Factor × Late Career | -0.000 | -0.000 | -0.002 | -0.000 | -0.000 |
| | (0.001) | (0.001) | (0.002) | (0.001) | (0.001) |
| Pre-Tenure | $0.11^{***}$ | $0.12^{***}$ | $0.09^{***}$ | $0.12^{***}$ | $0.09^{***}$ |
| | (0.016) | (0.017) | (0.023) | (0.017) | (0.023) |
| Tenure Eligible | 0.020 | 0.026 | -0.015 | 0.026 | -0.016 |
| | (0.013) | (0.014) | (0.019) | (0.014) | (0.019) |
| Mid-Career | 0.005 | 0.011 | -0.014 | 0.011 | -0.015 |
| | (0.013) | (0.013) | (0.018) | (0.013) | (0.018) |
| Late Career | 0.004 | 0.009 | -0.023 | 0.009 | -0.024 |
| | (0.012) | (0.012) | (0.016) | (0.012) | (0.016) |
| Constant | -0.261 | $-0.32^{*}$ | 0.055 | $-0.32^{*}$ | 0.060 |
| | (0.145) | (0.150) | (0.187) | (0.150) | (0.187) |
| Observations | 106,496 | 105,690 | 54,670 | 105,690 | 54,670 |
| F | 0.176 | 0.095 | 1.477 | 0.095 | 1.079 |
| *df* | 4.000 | 4.000 | 4.000 | 4.000 | 4.000 |
| p | 0.951 | 0.984 | 0.206 | 0.984 | 0.365 |

Notes. Experience is modeled using a linear spline with nodes at different career stages: pre-tenure teachers have between zero and two years of experience; tenure eligible teachers have between three and five years of experience; mid-career teachers have between six and ten years of experience; late career teachers have more than ten years of experience.

Robust standard errors in parentheses. $^{*}$ $p < 0.05$ $^{**}$ $p < 0.01$ $^{***}$ $p < 0.001$

Table A.3: Constructs, Survey Items, and Factor Loadings

| **Survey Item** | **Factor Loading** | **Standard Error** |
|---|---|---|
| *Panel A: Working Conditions and Mentoring* | | |
| 2018 | | |
| Working Conditions | | |
| TC_05c. The individual planning time provided for teachers in my school is sufficient. | 0.816 | 0.011 |
| TC_05d. The collaborative planning time provided for teachers in my school is sufficient. | 0.823 | 0.011 |
| Mentoring | | |
| TM_E04a. Focuses on the aspects of my work that will affect student learning | 0.914 | 0.002 |
| TM_E04b. Helps me to identify areas where my teaching is strong. | 0.919 | 0.002 |
| TM_E04c. Provides me with clear expectations for my teaching. | 0.939 | 0.002 |
| TM_E04d. Helps me plan instruction and develop lesson plans. | 0.874 | 0.003 |
| Working Conditions $\xleftrightarrow{\text{corr}}$ Mentoring | 0.384 | 0.01 |
| 2017 | | |
| Working Conditions | | |
| TC_3d. School leadership provides useful feedback about my instructional practices. | 0.91 | 0.004 |
| TC_4a. Teachers are encouraged to participate in school leadership roles. | 0.797 | 0.004 |
| TC_5b. The individual planning time provided for teachers in my school is sufficient. | 0.554 | 0.005 |
| TC_5c. The collaborative planning time provided for teachers in my school is sufficient. | 0.592 | 0.005 |
| Mentoring | | |
| TM_E4a. Focused on the aspects of my work that will affect student learning | 0.92 | 0.002 |
| TM_E4b. Helps me to identify areas where my teaching is strong. | 0.921 | 0.002 |

| | | |
|---|---|---|
| TM_E4c. Provides me with clear expectations for my teaching. | 0.94 | 0.001 |
| TM_E4d. Helps me plan instruction and develop lesson plans. | 0.887 | 0.002 |
| Working Conditions $\overset{\text{corr}}{\longleftrightarrow}$ Mentoring | 0.546 | 0.007 |
| 2016 | | |
| Working Conditions | | |
| TC_3d. School leadership provides useful feedback about my instructional practices. | 0.797 | 0.003 |
| TC_4a. Teachers have autonomy to make decisions about instruction e.g. pacing, materials, and pedagogy. | 0.713 | 0.003 |
| TC_4c. Teachers are encouraged to participate in school leadership roles. | 0.894 | 0.002 |
| TC_4d. Teachers have an appropriate level of influence on decision-making. | 0.899 | 0.002 |
| TC_4e. Teachers have opportunities to participate in shared leadership structures such as PLCs or grade and/or school level leadership teams. | 0.83 | 0.003 |
| TC_5b. The individual planning time provided for teachers in my school is sufficient. | 0.562 | 0.005 |
| TC_5c. The collaborative planning time provided for teachers in my school is sufficient. | 0.592 | 0.004 |
| Mentoring | | |
| TM_E8a. Focuses on the aspects of my work that will affect student learning. | 0.887 | 0.003 |
| TM_E8b. Helps me to identify areas where I can improve. | 0.943 | 0.002 |
| TM_E8c. Helps me to identify areas where my teaching is strong. | 0.922 | 0.002 |
| TM_E8d. Provides me with clear expectations for my teaching. | 0.92 | 0.002 |
| TM_E8e. Helps me plan instruction and develop lesson plans. | 0.903 | 0.003 |
| TM_E8f. Overall has helped refine my instructional practices. | 0.946 | 0.002 |

| | | |
|---|---|---|
| Working Conditions $\xleftrightarrow{\text{corr}}$ Mentoring | 0.52 | 0.012 |
| TC_5b $\xleftrightarrow{\text{corr}}$ TC_5c | 0.408 | 0.004 |
| 2015 | | |
| Working Conditions | | |
| Q4c. School leadership provides useful feedback about my instructional practices. | 0.837 | 0.003 |
| Q5a. Teachers have autonomy to make decisions about instruction e.g. pacing, materials, and pedagogy. | 0.705 | 0.004 |
| Q5b. Teachers are encouraged to participate in school leadership roles. | 0.819 | 0.003 |
| Q6b. The planning time provided for teachers in my school is sufficient. | 0.57 | 0.005 |
| Q6c. School leaders protect instructional time. | 0.742 | 0.004 |
| Mentoring | | |
| TM_E1a. Focuses on the aspects of my work that will affect student learning. | 0.884 | 0.003 |
| TM_E1b. Helps me to identify areas where I can improve. | 0.946 | 0.002 |
| TM_E1c. Helps me to identify areas where my teaching is strong. | 0.927 | 0.003 |
| TM_E1d. Provides me with clear expectations for my teaching. | 0.911 | 0.003 |
| TM_E1e. Helps me plan instruction and develop lesson plans. | 0.902 | 0.003 |
| TM_E1f. Overall has helped improve my instructional practices. | 0.933 | 0.002 |
| TM_E2a. Observed teaching your class? | 0.193 | 0.015 |
| TM_E2b. Given feedback on your teaching? | 0.275 | 0.014 |
| Working Conditions $\xleftrightarrow{\text{corr}}$ Mentoring | 0.545 | 0.011 |
| TM_E2a $\xleftrightarrow{\text{corr}}$ TM_E2b | 0.788 | 0.007 |
| TM_Q6b $\xleftrightarrow{\text{corr}}$ TM_Q6c | 0.175 | 0.005 |
| *Panel B: Professional Development and Evaluative Feedback* | | |
| 2018 | | |
| Professional Development | | |

| | | |
|---|---|---|
| TC_15a. Our school staff is a learning community in which ideas and suggestions for improvement are encouraged. | 0.755 | 0.003 |
| TC_15b. I receive specific professional learning suggestions that are tailored to my needs. | 0.877 | 0.002 |
| TC_15c. My professional learning is closely aligned to the evaluation system in place in my district. | 0.863 | 0.002 |
| TC_15d. My professional learning is closely aligned to the instructional materials I use in class. | 0.86 | 0.002 |
| TC_15e. In general, the professional learning I have received this year has led to improvements in my teaching. | 0.898 | 0.001 |
| TM_E07b. Time such as planning or release time to complete evaluation materials | 0.687 | 0.006 |
| TM_E07c. Materials such as guidelines to facilitate the process | 0.712 | 0.005 |
| TM_E07d. Access to staff with specific expertise such as instructional coaches within and/or outside of my school | 0.702 | 0.006 |
| TM_E05g. Indicators from the teacher observation rubric are often referenced in informal discussions between teachers. | 0.615 | 0.007 |
| TM_E05h. Indicators from the teacher observation rubric are often referenced in formal meetings where teaching is discussed. | 0.668 | 0.006 |
| Evaluative Feedback | | |
| TC_16a. The processes used to conduct my teacher evaluation are fair to me. | 0.714 | 0.004 |
| TC_16b. In general, the teacher evaluation process used in my school has led to improvements in my teaching. | 0.798 | 0.004 |
| TM_E05e. I receive very detailed feedback on my strengths and weaknesses through the evaluation process. | 0.873 | 0.004 |
| TM_E05f. The individuals who observe my classroom has the expertise to evaluate my practice. | 0.802 | 0.005 |

| | | |
|---|---|---|
| Learning $\xleftrightarrow{\text{corr}}$ Evaluation | 0.872 | 0.003 |
| TM_E05g $\xleftrightarrow{\text{corr}}$ TM_E05h | 0.378 | 0.006 |
| TC_16a $\xleftrightarrow{\text{corr}}$ TC_16b | 0.262 | 0.005 |
| TM_E07b $\xleftrightarrow{\text{corr}}$ TM_E07c | 0.392 | 0.007 |
| TM_E07c $\xleftrightarrow{\text{corr}}$ TM_E07d | 0.319 | 0.006 |
| TM_E07b $\xleftrightarrow{\text{corr}}$ TM_E07d | 0.292 | 0.006 |
| 2017 | | |
| Professional Development | | |
| TC_18a. I receive specific suggestions for professional learning that are tailored to my needs. | 0.726 | 0.004 |
| TC_18b. I receive professional development support that enhances my abilities to implement instructional strategies that meet the diverse needs of student learners. | 0.725 | 0.004 |
| TM_E7a. Leadership support such as key information and guidance from school or district administrators | 0.896 | 0.002 |
| TM_E7b. Time such as planning or release time to complete evaluation materials | 0.886 | 0.002 |
| TM_E7c. Materials such as guidelines to facilitate the process | 0.924 | 0.002 |
| TM_E7d. Access to staff with specific expertise such as instructional coaches within and/or outside of my school | 0.836 | 0.003 |
| Evaluative Feedback | | |
| TC_15c. The processes used to conduct my teacher evaluation are fair to me. | 0.742 | 0.004 |
| TC_15d. In general, the teacher evaluation process used in my school has led to improvements in my teaching. | 0.803 | 0.004 |
| TM_E5e. I receive very detailed feedback on my strengths and weaknesses through the evaluation process. | 0.828 | 0.004 |
| TM_E5f. The individuals who observe my classroom has the expertise to evaluate my practice. | 0.786 | 0.004 |

| | | |
|---|---|---|
| TM_E8a. TVAAS scores should be used to inform professional learning priorities. | 0.401 | 0.008 |
| TM_E8b. I find student growth data useful for organizing my classroom e.g. seating charts, etc. during beginning of year planning. | 0.415 | 0.008 |
| Learning $\xleftrightarrow{\text{corr}}$ Evaluation | 0.827 | 0.004 |
| TM_E08a $\xleftrightarrow{\text{corr}}$ TM_E08b | 0.485 | 0.006 |
| TC_18a $\xleftrightarrow{\text{corr}}$ TC_18b | 0.342 | 0.006 |
| TC_15c $\xleftrightarrow{\text{corr}}$ TC_15d | 0.196 | 0.005 |
| 2016 | | |
| Professional Development | | |
| TM_E10a. Leadership support such as key information and guidance from school or district administrators | 0.896 | 0.003 |
| TM_E10b. Time such as planning or release time to complete evaluation materials | 0.886 | 0.004 |
| TM_E10c. Materials such as guidelines to facilitate the process | 0.925 | 0.003 |
| TM_E10d. Access to staff with specific expertise such as instructional coaches within and/or outside of my school | 0.857 | 0.004 |
| Evaluative Feedback | | |
| TC_13a. Indicators from the teacher observation rubric are often referenced in informal discussions between teachers. | 0.573 | 0.007 |
| TC_13b. Indicators from the teacher observation rubric are often referenced in formal meetings where teaching is discussed. | 0.63 | 0.006 |
| TC_13c. The processes used to conduct my teacher evaluation are fair to me. | 0.738 | 0.006 |
| TC_13d. In general, the teacher evaluation process used in my school has led to improvements in my teaching. | 0.791 | 0.006 |
| TM_E9a. The system for assessing teachers as a whole generates fair and accurate results. | 0.906 | 0.004 |

| | | |
|---|---|---|
| TM_E9b. The evaluation criteria are applied equally to all teachers, regardless of their background or level of experience. | 0.813 | 0.005 |
| TM_E9c. The system is effective in identifying outstanding teachers. | 0.873 | 0.004 |
| Learning $\xleftrightarrow{\text{corr}}$ Evaluation | 0.672 | 0.009 |
| TC_13a $\xleftrightarrow{\text{corr}}$ TC_13b | 0.366 | 0.007 |
| TC_13c $\xleftrightarrow{\text{corr}}$ TC_13d | 0.203 | 0.007 |
| 2015 | | |
| Professional Development | | |
| Q11a. Indicators from the teacher observation rubric are often referenced in informal discussions between teachers. | 0.459 | 0.006 |
| Q11b. Indicators from the teacher observation rubric are often referenced in formal meetings where teaching is discussed. | 0.52 | 0.006 |
| Q11c. The processes used to conduct my teacher evaluation are fair to me. | 0.837 | 0.003 |
| Q11d. In general, the teacher evaluation process used in my school has led to improvements in my teaching. | 0.875 | 0.003 |
| TM_E8a. Leadership support such as key information and guidance from school or district administrators | 0.871 | 0.004 |
| TM_E8b. Time such as planning or release time to complete evaluation materials | 0.862 | 0.004 |
| TM_E8c. Materials such as guidelines to facilitate the process | 0.893 | 0.003 |
| TM_E8d. Access to staff with specific expertise such as instructional coaches within and/or outside of my school | 0.808 | 0.005 |
| Evaluative Feedback | | |
| TM_E7a. In my school, evaluation criteria and indicators are appropriate. | 0.888 | 0.004 |
| TM_E7b. The instruments used to measure teacher performance are easy to understand. | 0.835 | 0.005 |

| | | |
|---|---|---|
| TM_E7d. The evaluation criteria are applied equally to all teachers, regardless of their background or level of experience. | 0.826 | 0.005 |
| TM_E7e. The system is effective in identifying outstanding teachers. | 0.856 | 0.005 |
| TM_E7f. The system is effective in identifying teachers who are struggling. | 0.799 | 0.006 |
| Learning $\xleftrightarrow{\text{corr}}$ Evaluation | 0.778 | 0.006 |
| Q11a $\xleftrightarrow{\text{corr}}$ Q11b | 0.451 | 0.005 |
| TM_E07e $\xleftrightarrow{\text{corr}}$ TM_E07f | 0.185 | 0.006 |
| 2014 | | |
| Professional Development | | |
| Q64a. The processes used to conduct my teacher evaluation are fair to me. | 0.714 | 0.004 |
| Q64b. The teacher evaluation process causes me a lot of stress. | -0.339 | 0.006 |
| Q64c. The teacher evaluation process helps me improve as a professional. | 0.863 | 0.002 |
| Q64d. I made changes to my teaching based on my evaluation results. | 0.62 | 0.004 |
| Q64f. The process of evaluating my teaching performance takes more effort than the results are worth. | -0.578 | 0.004 |
| Q64g. The teacher evaluation process clearly defines what is expected of me. | 0.669 | 0.004 |
| Q64h. My observers are qualified to evaluate my teaching. | 0.581 | 0.005 |
| Q64j. In general, the teacher evaluation process used in my school will improve student learning | 0.895 | 0.002 |
| Q64k. Teaching observations disrupt my classroom instruction. | -0.468 | 0.005 |
| Q64l. Feedback from my teacher evaluation influences the professional development activities in which I participate. | 0.62 | 0.004 |
| Q64m. The teacher evaluation process used in my school will improve my teaching. | 0.954 | 0.001 |

| | | |
|---|---|---|
| Q64n. The teacher evaluation process used in my school will improve my students' achievement. | 0.932 | 0.001 |
| Q64o. I find it difficult to use feedback from my teaching observations to improve my practice. | -0.55 | 0.005 |
| Q64p. Overall, I am satisfied with the teacher evaluation process used in my school. | 0.837 | 0.002 |
| Evaluative Feedback | | |
| Q58a. My evaluator uses the rubrics from our teacher evaluation process as a basis for suggesting how I can improve my teaching. | 0.749 | 0.005 |
| Q58b. I use the rubric from our teacher evaluation process when planning instruction. | 0.708 | 0.005 |
| Q58c. Indicators from our teacher evaluation rubric are often referenced in formal meetings where teaching is discussed. | 0.853 | 0.004 |
| Q58d. Indicators from our teacher evaluation rubric are often referenced in informal discussions between teachers. | 0.786 | 0.004 |
| Learning $\xleftrightarrow{\text{corr}}$ Evaluation | 0.481 | 0.006 |
| Q64a $\xleftrightarrow{\text{corr}}$ Q64p | 0.167 | 0.003 |
| Q64b $\xleftrightarrow{\text{corr}}$ Q64f | 0.27 | 0.005 |
| Q58a $\xleftrightarrow{\text{corr}}$ Q64h | 0.261 | 0.005 |
| 2013 | | |
| Professional Development | | |
| Q43a. The processes used to conduct my teacher evaluation are fair to me. | 0.774 | 0.003 |
| Q43c. The teacher evaluation process helps me improve as a professional. | 0.836 | 0.003 |
| Q43d. The process of evaluating my teaching performance takes more effort than the results are worth. | -0.556 | 0.005 |
| Q43h. Feedback from my teacher evaluation influences the professional development activities in which I participate. | 0.61 | 0.005 |
| Q43j. The teacher evaluation process used in my school will improve my students' achievement. | 0.838 | 0.003 |

| | | |
|---|---|---|
| Q43k. Overall, I am satisfied with the teacher evaluation process used in my school. | 0.899 | 0.002 |
| Q35a. The specific indicators of teaching performance in the rubric(s) used in my school's teacher evaluation process accurately reflect what teachers know and do. | 0.747 | 0.004 |
| Q35d. I believe I can achieve the highest rating on most elements of teaching performance defined in the rubric(s) used in my school's teacher evaluation process. | 0.513 | 0.005 |
| Q35e. Teachers must receive a score of 4 or higher on all indicators on the rubric(s) used for teaching observations to be rated as an effective teacher. | 0.255 | 0.007 |
| Q35f. The rubric(s) used in my school's teacher evaluation process clearly describe the teaching performance needed to earn each rating score. | 0.654 | 0.005 |
| Q35g. The rubric(s) omit important aspects of teaching that should be considered when evaluating teachers. | -0.39 | 0.006 |
| Evaluative Feedback | | |
| Q35b. My evaluator uses the rubrics from our teacher evaluation process as a basis for discussing feedback from teaching observations. | 0.875 | 0.004 |
| Q35c. My evaluator uses the rubrics from our teacher evaluation process as a basis for suggesting how I can improve my teaching. | 0.987 | 0.004 |
| Learning $\xleftrightarrow{\text{corr}}$ Evaluation | 0.588 | 0.005 |
| 2012 | | |
| Professional Development | | |
| Q32a. I can accurately describe to others the processes and procedures used to conduct my teacher evaluation. | 0.555 | 0.006 |
| Q32c. The teacher evaluation process causes me a lot of stress. | -0.338 | 0.009 |
| Q32d. The teacher evaluation process helps me improve as a professional. | 0.692 | 0.006 |

| | | |
|---|---|---|
| Q32h. The teacher evaluation process clearly defines what is expected of me. | 0.78 | 0.004 |
| Q32k. Teaching observations disrupts classroom instruction. | -0.342 | 0.008 |
| Q30a. The specific indicators of teaching performance in the rubric(s) used in my school's teacher evaluation process accurately reflect what teachers know and do. | 0.755 | 0.005 |
| Q30d. Generally, each of my teaching observations has focused on just a portion or subset of the entire teacher evaluation rubric. | 0.201 | 0.009 |
| Q30e. I believe I can achieve the highest rating on most elements of teaching performance defined in the rubric(s) used in my school's teacher evaluation process. | 0.512 | 0.007 |
| Q30f. The rubric(s) used in my school's teacher evaluation process clearly describe the teaching performance needed to earn each rating score. | 0.783 | 0.004 |
| Q30g. The rubric(s) omit important aspects of teaching that should be considered when evaluating teachers. | -0.386 | 0.007 |
| Evaluative Feedback | | |
| Q30b. My evaluator uses the rubric(s) from our teacher evaluation process as a basis for discussing feedback from teaching observations. | 0.835 | 0.006 |
| Q30c. My evaluator uses the rubrics from our teacher evaluation process as a basis for suggesting how I can improve my teaching. | 0.967 | 0.006 |
| Q32c $\xleftrightarrow{\text{corr}}$ Q32k | 0.279 | 0.009 |

# Bibliography


Abelson, M. A. (1987). Examination of avoidable and unavoidable turnover. *Journal of Applied Psychology*, *72*(3), 382. https://doi.org/10.1037/0021-9010.72.3.382

Allen, R., & Sims, S. (2017). *Improving science teacher retention: Do national STEM learning network professional development courses keep science teachers in the classroom*, Wellcome Trust. https://www.stem.org.uk/system/files/elibrary-resources/2017/09/Improving%20Science%20Teacher%20Retention.pdf

Arellano, M., & Hahn, J. (2007). Understanding bias in nonlinear panel models: Some recent developments. *Econometric Society Monographs*, *43*, 381.

Auletto, A. (2021). Making sense of early-career teacher support, satisfaction, and commitment. *Teaching and Teacher Education*, *Online*. https://doi.org/10.1016/j.tate.2021.103321

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. https://doi.org/10.1177/0022487108324554

Bardelli, E., Ronfeldt, M., & Papay, J. (2021, August). *Teacher preparation programs and graduates' growth in instructional effectiveness* (EdWorkingPapers ai21-450), Annenberg Institute for School Reform at Brown University. https://doi.org/10.26300/120j-qb52

Barrett, C. A., & Pas, E. T. (2020). A cost analysis of traditional professional development and coaching structures in schools. *Prevention Science*, *21*(5), 604–614. https://doi.org/10.1007/s11121-020-01115-5

Bollen, K. A. (1989). *Structural equations with latent variables*. John Wiley & Sons, Inc. https://doi.org/10.1002/9781118619179

Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher*, *33*(8), 3–15. https://doi.org/10.3102/0013189x033008003

Boyd, D., Lankford, H., Loeb, S., Ronfeldt, M., & Wyckoff, J. (2010). The role of teacher quality in retention and hiring: Using applications to transfer to uncover preferences of teachers and schools: The role of teacher quality in retention and hiring. *Journal of Policy Analysis and Management*, *30*(1), 88–110. https://doi.org/10.1002/pam.20545

Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2008). *Who leaves? teacher attrition and student achievement* (Working Paper No. 14022). National Bureau of Economic Research. https://doi.org/10.3386/w14022

Bruno, P., Rabovsky, S. J., & Strunk, K. O. (2019). Taking their first steps: The distribution of new teachers in school and classroom contexts and implications for teacher effectiveness. *American Educational Research Journal*, *57*(4), 1688–1729. https://doi.org/10.3102/0002831219882008

Bryk, A. S., & Schneider, B. (2002). *Trust in schools: A core resource for improvement*. Russell Sage Foundation.

Bryk, A. S., Sebring, P. B., Allensworth, E., Easton, J. Q., & Luppescu, S. (2009). *Organizing schools for improvement: Lessons from chicago*. University of Chicago Press. https://doi.org/10.7208/chicago/9780226078014.001.0001

Burgess, S., Rawal, S., & Taylor, E. S. (2019). *Teacher peer observation and student test scores: Evidence from a field experiment in English secondary schools* (EdWorkingPaper No. 19-139), Annenberg Institute at Brown University. https://doi.org/10.26300/ry5b-g146

Cohen, D. K. (2011, August 31). *Teaching and its predicaments*. Harvard University Press.

Cohen, D. K., Raudenbush, S. W., & Ball, D. L. (2003). Resources, instruction, and research. *Educational Evaluation and Policy Analysis*, *25*(2), 119–142. https://doi.org/10.3102/01623737025002119

Cullen, J. B., Koedel, C., & Parsons, E. (2021). The compositional effect of rigorous teacher evaluation on workforce quality. *Education Finance and Policy*, *16*(1), 7–41. https://doi.org/10.1162/edfp_a_00292

Darling-Hammond, L. (1997). *Doing what matters most: Investing in quality teaching*, National Commission on Teaching & America's Future. https://eric.ed.gov/?id=ED415183

Davis, B., & Higdon, K. (2008). The effects of Mentoring/Induction support on beginning teachers' practices in early elementary classrooms (K-3). *Journal of Research in Childhood Education*, *22*(3), 261–274. https://doi.org/10.1080/02568540809594626

Dee, T. S., & Wyckoff, J. (2015). Incentives, selection, and teacher performance: Evidence from IMPACT: Incentives, selection, and teacher performance. *Journal of Policy Analysis and Management*, *34*(2), 267–297. https://doi.org/10.1002/pam.21818

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, *38*(3), 181–199. https://doi.org/10.3102/0013189x08331140

Desimone, L. M., Hochberg, E. D., & Mcmaken, J. (2016). Teacher knowledge and instructional quality of beginning teachers: Growth and linkages. *Teachers College Record*, *118*(5), 1–54. https://doi.org/10.1177/016146811611800507

Desimone, L. M., Hochberg, E. D., Porter, A. C., Polikoff, M. S., Schwartz, R., & Johnson, L. J. (2013). Formal and informal mentoring: Complementary, compensatory, or consistent? *Journal of Teacher Education*, *65*(2), 88–110. https://doi.org/10.1177/0022487113511643

Erickson, S. J. (2007). *An examination of the relationship between professional development and teacher turnover* (Doctoral dissertation). University of Oregon. https://www.proquest.com/docview/304819862

Evertson, C. M., & Smithey, M. W. (2000). Mentoring effects on protégés' classroom practice: An experimental field study. *The Journal of Educational Research*, *93*(5), 294–304. https://doi.org/10.1080/00220670009598721

Fan, X., & Sivo, S. A. (2007). Sensitivity of fit indices to model misspecification and model types. *Multivariate Behavioral Research*, *42*(3), 509–529. https://doi.org/10.1080/00273170701382864

Fernández-Val, I. (2009). Fixed effects estimation of structural parameters and marginal effects in panel probit models. *Journal of Econometrics*, *150*(1), 71–85.

Fletcher, S. H., & Strong, M. A. (2009). Full-release and site-based mentoring of new elementary grade teachers: An analysis of changes in student achievement. *The New Educator*, *5*(4), 329–341. https://doi.org/10.1080/1547688x.2009.10399583

Fletcher, S. H., Strong, M. A., & Villar, A. (2008). An investigation of the effects of variations in mentor-based induction on the performance of students in California. *Teachers College Record: The Voice of Scholarship in Education*, *110*(10), 2271–2289. https://doi.org/10.1177/016146810811001003

Geiger, T., & Pivovarova, M. (2018). The effects of working conditions on teacher retention. *Teachers and Teaching*, *24*(6), 604–625. https://doi.org/10.1080/13540602.2018.1457524

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press.

Glazerman, S., Isenberg, E., Dolfin, S., Bleeker, M., Johnson, A., Grider, M., & Jacobus, M. (2010). *Impacts of comprehensive teacher induction: Final results from a randomized controlled study* (NCEE 2010-4027), U.S. Department of Education. https://eric.ed.gov/?id=ED565837

Goldhaber, D., Gross, B., & Player, D. (2010). Teacher career paths, teacher quality, and persistence in the classroom: Are public schools keeping their best? Are public schools keeping their best? *Journal of Policy Analysis and Management*, *30*(1), 57–87. https://doi.org/10.1002/pam.20549

Greenberg, D., & McCall, J. (1974). Teacher mobility and allocation. *The Journal of Human Resources*, *9*(4), 480. https://doi.org/10/d7f5x8

Grissom, J. A., & Bartanen, B. (2018). Strategic retention: Principal effectiveness and teacher turnover in multiple-measure teacher evaluation systems. *American Educational Research Journal*, *56*(2), 514–555. https://doi.org/10.3102/0002831218797931

Grissom, J. A., & Loeb, S. (2011). Triangulating principal effectiveness: How perspectives of parents, teachers, and assistant principals identify the central importance of managerial skills. *American Educational Research Journal*, *48*(5), 1091–1123. https://doi.org/10.3102/0002831211402663

Guarino, C. M., Santibañez, L., & Daley, G. A. (2006). Teacher recruitment and retention: A review of the recent empirical literature. *Review of Educational Research*, *76*(2), 173–208. https://doi.org/10.3102/00346543076002173

Haig, B. D. (2018). *Exploratory factor analysis*. Oxford University Press. https://doi.org/10.1093/oso/9780190222055.003.0006

Hammerness, K., & Matsko, K. K. (2012). When context has content: A case study of new teacher induction in the university of chicago's urban teacher education program. *Urban Education*, *48*(4), 557–584. https://doi.org/10.1177/0042085912456848

Hanushek, E. (2009). Teacher deselection. In D. Goldhaber & J. Hannaway (Eds.), *Creating a new teaching profession* (pp. 165–180). Urban Institute Press.

Hanushek, E., & Rivkin, S. (2010). *Constrained job matching: Does teacher job search harm disadvantaged urban schools?* (Working Paper No. 15816). National Bureau of Economic Research. https://doi.org/10.3386/w15816

Hanushek, E., Rivkin, S., & Schiman, J. (2016). *Dynamic effects of teacher turnover on the quality of instruction* (Working Paper No. 22472). National Bureau of Economic Research. National Bureau of Economic Research. https://doi.org/10.3386/w22472

Henke, R. R., Chen, X., & Geis, S. (2000). *Progress through the teacher pipeline: 1992-93 college graduates and Elementary/Secondary school teaching as of 1997* (NCES 2000–152), U.S. Department of Education. National Center for Education Statistics. https://eric.ed.gov/?id=ED438252

Henry, G. T., & Redding, C. (2020). The consequences of leaving school early: The effects of within-year and end-of-year teacher turnover. *Education Finance and Policy*, *15*(2), 332–356. https://doi.org/10.1162/edfp_a_00274

Herbst, P., & Chazan, D. (2016). The role of theory development in increasing the subject specificity of research on mathematics teaching. In J. Cai (Ed.), *Compendium for research in mathematics education* (pp. 102–127).

Horn, I. S. (2010). Teaching replays, teaching rehearsals, and re-visions of practice: Learning from colleagues in a mathematics teacher community. *Teachers College Record*, *112*(1), 225–259. https://doi.org/10.1177/016146811011200109

Howe, E. R. (2006). Exemplary teacher induction: An international review. *Educational Philosophy and Theory*, *38*(3), 287–297. https://doi.org/10.1111/j.1469-5812.2006.00195.x

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Ingersoll, R. M., & Merrill, L. (2012). Seven trends: The transformation of the teaching force: (579212012-001). https://doi.org/10.1037/e579212012-001

Ingersoll, R. M., Merrill, L., & May, H. (2014). *What are the effects of teacher education and preparation on beginning teacher attrition?*, CPRE Research Reports. https://repository.upenn.edu/cpre_researchreports/78

Ingersoll, R. M., & Smith, T. M. (2004). Do teacher induction and mentoring matter? *NASSP Bulletin*, *88*(638), 28–40. https://doi.org/10.1177/019263650408863803

Ingersoll, R. M., & Strong, M. (2011). The impact of induction and mentoring programs for beginning teachers: A critical review of the research. *Review of Educational Research*, *81*(2), 201–233. https://doi.org/10.3102/0034654311403323

Isenberg, E., Glazerman, S., Bleeker, M., Johnson, A., Lugo-Gil, J., Grider, M., Dolfin, S., & Britton, E. (2009). *Impacts of comprehensive teacher induction: Results from the second year of a randomized controlled study* (NCEE 2009-4072), National Center for Education Evaluation and Regional Assistance. https://eric.ed.gov/?id=ED511788

Jackson, C. K., & Bruegmann, E. (2009). Teaching students and teaching each other: The importance of peer learning for teachers. *American Economic Journal: Applied Economics*, *1*(4), 85–108. https://doi.org/10.1257/app.1.4.85

Kang, S., & Berliner, D. C. (2012). Characteristics of teacher induction programs and turnover rates of beginning teachers. *The teacher educator*, *47*(4), 268–282. https://doi.org/10.1080/08878730.2012.707758

Kline, P. (2014). *An easy guide to factor analysis*. Routledge. https://doi.org/10.4324/9781315788135

Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford Publications.

Koedel, C., Li, J., Springer, M. G., & Tan, L. (2017). The impact of performance ratings on job satisfaction for public school teachers. *American Educational Research Journal*, *54*(2), 241–278. https://doi.org/10.3102/0002831216687531

Kraft, M. A., Blazar, D., & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. *Review of educational research*, *88*(4), 547–588. https://doi.org/10.3102/0034654318759268

Kraft, M. A., & Hill, H. C. (2020). Developing ambitious mathematics instruction through web-based coaching: A randomized field trial. *American Educational Research Journal*, *57*(6), 2378–2414. https://doi.org/10.3102/0002831220916840

Kraft, M. A., & Papay, J. P. (2014). Can professional environments in schools promote teacher development? explaining heterogeneity in returns to teaching experience. *Educational Evaluation and Policy Analysis*, *36*(4), 476–500. https://doi.org/10.3102/0162373713519496

Ladd, H. F. (2011). Teachers' perceptions of their working conditions: How predictive of planned and actual teacher movement? *Educational Evaluation and Policy Analysis*, *33*(2), 235–261. https://doi.org/10.3102/0162373711398128

Little, J. W. (1987). Teachers as colleagues. In V. Richardson-Koehler (Ed.), *Educators' handbook: A research perspective* (pp. 491–518).

Loeb, S., Darling-Hammond, L., & Luczak, J. (2005). How teaching conditions predict teacher turnover in California schools. *Peabody Journal of Education*, *80*(3), 44–70. https://doi.org/10.1207/s15327930pje8003_4

Lortie, D. C. (1975). *School-teacher: A sociological study*. The University of Chicago Press.

Luft, J., & Cox, W. (2001). Investing in our future: A survey of support offered to beginning secondary science and mathematics teachers. *Science Educator*, *10*(1), 1–9.

Lynch, K., Hill, H. C., Gonzalez, K. E., & Pollard, C. (2019). Strengthening the research base that informs STEM instructional improvement efforts: A meta-analysis. *Educational Evaluation and Policy Analysis*, 0162373719849044. https://doi.org/10.3102/0162373719849044

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin*, *103*(3), 391–410. https://doi.org/10.1037/0033-2909.103.3.391

Matsko, K. K., Coca, V., & Easton, J. (2007). *Keeping new teachers: A first look at the influences of induction in the Chicago public schools*, Consortium on Chicago School Research.

Maydeu-Olivares, A. (2017). Assessing the size of model misfit in structural equation models. *Psychometrika*, *82*(3), 533–558. https://doi.org/10.1007/s11336-016-9552-7

Miller, J. M., Youngs, P., Perrone, F., & Grogan, E. (2020). Using measures of fit to predict beginning teacher retention. *The Elementary School Journal*, *120*(3), 399–421. https://doi.org/10.1086/707094

Moore Johnson, S. (2006). *The workplace matters: Teacher quality, retention, and effectiveness*, National Education Association. https://eric.ed.gov/?id=ED495822

Moore Johnson, S. (2020). *Where teachers thrive: Organizing schools for success*. Harvard Education Press.

Moore Johnson, S., Kraft, M. A., & Papay, J. P. (2012). How context matters in high-need schools: The effects of teachers' working conditions on their professional satisfaction and their students' achievement. *Teachers College Record*, *114*(10), 1–39. https://doi.org/10.1177/016146811211401004

Murnane, R. J. (1981). Teacher mobility revisited. *The Journal of Human Resources*, *16*(1), 3. https://doi.org/10.2307/145216

Muthen, B. O., Muthen, L., & Asparouhov, T. (2015). *Estimator choices with categorical outcomes*. https://www.statmodel.com/download/EstimatorChoices.pdf

Nevins Stanulis, R., & Floden, R. E. (2009). Intensive mentoring as a way to help beginning teachers develop balanced instruction. *Journal of Teacher Education*, *60*(2), 112–122. https://doi.org/10.1177/0022487108330553

Nguyen, T. D., Pham, L., Springer, M., & Crouch, M. (2019). *The factors of teacher attrition and retention: An updated and expanded meta-analysis of the literature* (EdWorkingPapers No. 19-149). Annenberg Institute at Brown University. https://doi.org/10.26300/cdf3-4555

Papay, J. P. (2012). Refocusing the debate: Assessing the purposes and tools of teacher evaluation. *Harvard Educational Review*, *82*(1), 123–141. https://doi.org/10.17763/haer.82.1.v40p0833345w6384

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. https://doi.org/10.1016/j.jpubeco.2015.02.008

Papay, J. P., Taylor, E. S., Tyler, J. H., & Laski, M. E. (2020). Learning job skills from colleagues at work: Evidence from a field experiment using teacher performance data. *American Economic Journal: Economic Policy*, *12*(1), 359–388. https://doi.org/10.1257/pol.20170709

Phelps, G., Kelcey, B., Jones, N., & Liu, S. (2016). Informing estimates of program effects for studies of mathematics professional development using teacher content knowledge outcomes. *Evaluation Review*, *40*(5), 383–409. https://doi.org/10.1177/0193841x16665024

Preacher, K. J., & MacCallum, R. C. (2003). Repairing tom swift's electric factor analysis machine. *Understanding Statistics*, *2*(1), 13–43. https://doi.org/10.1207/s15328031us0201_02

Rabe-Hesketh, S., & Skrondal, A. (2012). *Multilevel and longitudinal modeling using stata*. STATA press.

Raykov, T. (1997). Estimation of composite reliability for congeneric measures. *Applied Psychological Measurement*, *21*(2), 173–184. https://doi.org/10.1177/01466216970212006

Redding, C., Booker, L. N., Smith, T. M., & Desimone, L. M. (2019). School administrators' direct and indirect influences on middle school math teachers' turnover. *Journal of Educational Administration*, *57*(6), 708–730. https://doi.org/10.1108/jea-10-2018-0190

Rockoff, J. (2008). *Does mentoring reduce turnover and improve skills of new employees? evidence from teachers in new York city* (Working Paper No. 13868). National Bureau of Economic Research. https://doi.org/10.3386/w13868

Rodriguez, L. A., Swain, W. A., & Springer, M. G. (2020). Sorting through performance evaluations: The influence of performance evaluation reform on teacher attrition and mobility. *American Educational Research Journal*, *57*(6), 2339–2377. https://doi.org/10.3102/0002831220910989

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. https://doi.org/10.3102/0162373711420865

Ronfeldt, M., Farmer, S. O., McQueen, K., & Grissom, J. A. (2015). Teacher collaboration in instructional teams and student achievement. *American Educational Research Journal*, *52*(3), 475–514. https://doi.org/10.3102/0002831215585562

Ronfeldt, M., Loeb, S., & Wyckoff, J. (2013). How teacher turnover harms student achievement. *American Educational Research Journal*, *50*(1), 4–36. https://doi.org/10.3102/0002831212463813

Ronfeldt, M., & McQueen, K. (2017). Does new teacher induction really improve retention? *Journal of Teacher Education*, *68*(4), 394–410. https://doi.org/10.1177/0022487117702583

Sackett, D. L., & Haynes, R. B. (Eds.). (1997). Summarising the effects of therapy: A new table and some more terms. *ACP Journal Club*, *127*(1), A15. https://doi.org/10.7326/acpjc-1997-127-1-a15

Sivo, S. A., Fan, X., Witta, E. L., & Willse, J. T. (2006). The search for ”Optimal” cutoff properties: Fit index criteria in structural equation modeling. *The Journal of Experimental Education*, *74*(3), 267–288. https://doi.org/10.3200/jexe.74.3.267-288

Smith, T. M., Booker, L. N., Hochberg, E. D., & Desimone, L. M. (2018). Do organizational supports for math instruction improve the quality of beginning teachers' instruction? *Teachers College Record: The Voice of Scholarship in Education*, *120*(7), 1–46. https://doi.org/10.1177/016146811812000702

Smith, T. M., & Ingersoll, R. M. (2004). What are the effects of induction and mentoring on beginning teacher turnover? *American Educational Research Journal*, *41*(3), 681–714. https://doi.org/10.3102/00028312041003681

Snijders, T. A., & Bosker, R. J. (2011). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. sage.

Song, M., Wayne, A. J., Garet, M. S., Brown, S., & Rickles, J. (2021). Impact of providing teachers and principals with performance feedback on their practice and student achievement: Evidence from a large-scale randomized experiment. *National Center for Education Evaluation and Regional Assistance*, *14*(2), 353–378. https://doi.org/10.1080/19345747.2020.1868030

Sorensen, L. C., & Ladd, H. F. (2020). The hidden costs of teacher turnover. *AERA Open*, *6*(1), 233285842090581. https://doi.org/10.1177/2332858420905812

Steinberg, M. P., & Garrett, R. (2016). Classroom composition and measured teacher performance: What do teacher observation scores really measure? *Educational Evaluation and Policy Analysis*, *38*(2), 293–317. https://doi.org/10.3102/0162373715616249

Steinberg, M. P., & Sartain, L. (2015). Does teacher evaluation improve school performance? experimental evidence from Chicago's excellence in teaching project. *Education Finance and Policy*, *10*(4), 535–572. https://doi.org/10.1162/edfp_a_00173

Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, *80*(1), 99–103. https://doi.org/10.1207/s15327752jpa8001_18

Sun, M., Loeb, S., & Grissom, J. A. (2016). Building teacher teams: Evidence of positive spillovers from more effective colleagues. *Educational Evaluation and Policy Analysis*, *39*(1), 104–125. https://doi.org/10.3102/0162373716665698

Sutcher, L., Darling-Hammond, L., & Carver-Thomas, D. (2016a). *A coming crisis in teaching? teacher supply, demand, and shortages in the U.S.*, Learning Policy Institute. https://doi.org/10.54300/247.242

Sutcher, L., Darling-Hammond, L., & Carver-Thomas, D. (2016b). *A coming crisis in teaching? teacher supply, demand, and shortages in the U.S.*, Learning Policy Institute. https://doi.org/10.54300/247.242

Sutcher, L., Darling-Hammond, L., & Carver-Thomas, D. (2019). Understanding teacher shortages: An analysis of teacher supply and demand in the united states. *Educational Policy Analysis Archives*, *27*, 35. https://doi.org/10.14507/epaa.27.3696

Taylor, E. S., & Tyler, J. H. (2012). The effect of evaluation on teacher performance. *American Economic Review*, *102*(7), 3628–3651. https://doi.org/10.1257/aer.102.7.3628

Thompson, M., Paek, P., Goe, L., & Ponte, E. (2005). The impact of new teacher induction on teacher practices and student learning. *Paper presented at the Annual Meeting of the American Educational Research Association*. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.568.1810&rep=rep1&type=pdf

U.S. Department of Education. (2009). *Race to the top: Executive summary*. https://eric.ed.gov/?id=ED557422

Watlington, E., Shockley, R., Guglielmino, P., & Felsher, R. (2010). The high cost of leaving: An analysis of the cost of teacher turnover. *Journal of Education Finance*, *36*(1), 22–37. https://doi.org/10.1353/jef.0.0028

Wechsler, M. E., Caspary, K., Humphrey, D. C., & Matsko, K. K. (2012). Examining the effects of new teacher induction. *Teachers College Record*, *114*(14), 387–416. https://doi.org/10.1177/016146811211401408

Worthington, R. L., & Whittaker, T. A. (2006). Scale development research: A content analysis and recommendations for best practices. *Counseling Psychologist*, *34*(6), 806–838. https://doi.org/10.1177/0011000006288127

Ximénez, C., Maydeu-Olivares, A., Shi, D., & Revuelta, J. (2022). Assessing cutoff values of SEM fit indices: Advantages of the unbiased SRMR index and its cutoff criterion based on communality. *Structural Equation Modeling: A Multidisciplinary Journal*, *29*(3), 368–380. https://doi.org/10.1080/10705511.2021.1992596

Yoon, K. S., Duncan, T., Lee, S. W.-Y., Scarloss, B., & Shapley, K. L. (2007). *Reviewing the evidence on how teacher professional development affects student achievement. issues & answers* (REL 2007-No. 033), Regional Educational Laboratory Southwest (NJ1). http://eric.ed.gov/?id=ED498548

Zeichner, K. M., & Tabachnick, B. R. (1981). Are the effects of university teacher education “washed out” by school experience? *Journal of Teacher Education*, *32*(3), 7–11. https://doi.org/10.1177/002248718103200302