AAAS | ARISE

Advancing Research & Innovation in the STEM Education of Preservice Teachers in High-Need School Districts

Teacher Induction Programs that Lead to Retention in the STEM Teaching Workforce

Peter Youngs
Kristen Bieda
Jihyun Kim

A 2019 AAAS ARISE Commissioned Paper

Prepared for the
American Association for the Advancement of Science

Teacher Induction Programs that Lead to Retention in the STEM Teaching Workforce

Part of the 2019 AAAS ARISE Commissioned Paper Series
Prepared for the American Association for the Advancement of Science

Peter Youngs
University of Virginia

Kristen Bieda
Michigan State University

Jihyun Kim
University of Virginia

Contents

Acknowledgements & 2019 AAAS ARISE Commissioned Paper Series……………….………ii

Abstract ……………………………………………………………………………………..iii

Introduction…………………………………………………………………………………..1

Formal Mentoring and Induction Programs and STEM Teacher Retention, Instructional Quality, and Effectiveness……………………………………………………………………………….3

Mentoring, Induction, and Beginning STEM Teacher Retention…………………………4

Mentoring, Induction, and Beginning Teacher Instructional Quality……………………..8

Beginning Teachers' Instructional Quality Across Content Areas………………..9

Beginning STEM Teachers' Instructional Quality………………………………12

Other Research on Induction for Secondary STEM Teachers…………………...14

Mentoring, Induction, and Beginning Teacher Effectiveness……………………………15

Summary……………………………………………………………………………16

Principal Leadership and STEM Teacher Retention and Effectiveness…………………………18

Principal Leadership and Teacher Retention…………………………………………….18

Principal Leadership and STEM Teacher Effectiveness…………………………………22

Summary……………………………………………………………………………24

School Organizational Context, Commitment, Retention, and Instructional Quality………..…24

Summary……………………………………………………………………………26

Strength and Weaknesses of Methods and Indicators Used in Research on Beginning Teacher Induction Programs………………………………………………………………………26

Strengths……………………………………………………………………………27

Weaknesses…………………………………………………………………………………28

Implications for Future Research…………………………………………………...…29

New Methods and Indicators…………………………………………………….…29

Challenges to Research Implementation………………………………………………………30

Strengthening Linkages Among STEM Induction Researchers………………………………….33

Research on Mentoring and Induction Programs……………………………………...33

Research on Instructional Coaching……………………..………………………………34

Research on Principal Leadership and School Organizational Context………………….35

Common Research Methods and Data Collection Guidelines…………………………...35

References…………………………………………………………………………………..37

Appendix……………………………………………………………………………………41

Acknowledgements

This research was supported by the National Science Foundation (NSF) under Grant No. DUE-1548986 to the American Association for the Advancement of Science (AAAS). Any opinions, findings, interpretations, conclusions or recommendations expressed in this material are those of its authors and do not represent the views of the AAAS Board of Directors, the Council of AAAS, AAAS' membership or the National Science Foundation.

2019 AAAS ARISE Commissioned Paper Series

Wilson, S. M. (2019). *Introduction to 2019 AAAS ARISE Commissioned Paper Series*. American Association for the Advancement of Science.

1. Bell, C., Gitomer, D., Savage, C., & Mckenna, A. H. (2019). *A Synthesis of Research on and Measurement of STEM Teacher Preparation*. American Association for the Advancement of Science.

2. Fuller, E. J., & Pendola, A. (2019). *Teacher Preparation and Teacher Retention: Examining the Relationship for Beginning STEM Teachers*. American Association for the Advancement of Science.

3. Youngs, P., Bieda, K., & Kim, J. (2019). *Teacher Induction Programs Associated with Retention in the STEM Teaching Workforce*. American Association for the Advancement of Science.

Read the full papers at: https://aaas-arise.org/commissioned-papers/.

Suggested Citation

Youngs, P., Bieda, K., & Kim, J. (2019). *Teacher Induction Programs Associated with Retention in the STEM Teaching Workforce*. American Association for the Advancement of Science.

Abstract

STEM teacher retention is critical for improving STEM learning outcomes in the United States for several reasons. First, STEM teachers who remain in the profession generally become much more effective over time (Harris & Sass, 2011; Papay & Kraft, 2015). Second, students perform better in mathematics and other subjects when teacher turnover in their grade levels at their schools is reduced (Atteberry, Loeb, & Wyckoff, 2017; Ronfeldt, Loeb, & Wyckoff, 2013). Third, STEM teachers who remain in teaching and in their schools of origin are likely to contribute to a positive climate in their schools; in other words, schools that have high rates of teacher retention are more likely than other schools to establish and maintain supportive professional environments. Schools with supportive professional environments have higher levels of student achievement gains over time than other schools (Bryk & Schneider, 2002; Newmann, Smith, Allensworth, & Bryk, 2001). To this end, this essay examines what it known and what remains to be learned about school working conditions that promote novice STEM teacher retention. In the first three sections, we review research on how formal mentoring and induction programs, principal leadership, and person-environment fit affect beginning STEM teacher retention and two related outcomes, instructional quality and effectiveness. The fourth section both identifies strengths and weaknesses of the methods and indicators used for research on mentoring/induction programs, principal leadership, and teacher fit and STEM teachers; and recommends new methods and indicators for research in this area. In the fifth section, we discuss challenges to research implementation in this area and we conclude with several suggestions for strengthening linkages among STEM teacher induction scholars. These include identifying models that will permit cross-comparison of findings across studies, using common research methods, and creating data collection guidelines.

Introduction

Teacher retention is critical for improving learning outcomes in the United States for several reasons. First, teachers who remain in the profession generally become much more effective over time in mathematics and reading (Harris & Sass, 2011; Papay & Kraft, 2015). Second, students perform better in mathematics and English language arts (ELA) when teacher turnover in their grade levels at their schools is reduced (Atteberry, Loeb, & Wyckoff, 2017; Ronfeldt, Loeb, & Wyckoff, 2013). Third, teachers who remain in teaching and in their schools of origin are likely to contribute to a positive climate in their schools; in other words, schools with high rates of teacher retention are more likely than other schools to establish and maintain supportive professional environments. Schools with supportive professional environments have higher levels of student achievement gains over time in mathematics and reading than other schools (Bryk & Schneider, 2002; Newmann, Smith, Allensworth, & Bryk, 2001).

Research indicates that teachers tend to leave schools that serve students most under-represented in the STEM career pipeline, namely low-income students, racial/ethnic minority students, and/or low-achieving students, at higher rates than they leave other schools (Allensworth, Ponisciak, & Mazzeo, 2009; Lankford, Loeb, & Wyckoff, 2002; Scafidi, Sjoquist, & Stinebrickner, 2007). This finding applies to mathematics and science teachers (Ingersoll & May, 2012). Also, the schools that teachers leave tend to be ones with lower resources than other schools. However, school characteristics, including student composition, often become less important for teacher turnover when other school context variables, such as mentoring/induction programs, principal leadership, and person-environment fit are included in analyses (Johnson, Kraft, & Papay, 2012; Kim, 2017; Player, Youngs, Perrone, & Grogan, 2017; Smith & Ingersoll, 2004). In other words, beginning teachers are more likely to remain in their schools of origin when they experience supportive professional environments.

This essay examines what it known and what remains to be learned about formal mentoring and induction programs and other school working conditions that promote novice STEM teacher retention. In this essay, we define *STEM teachers* as (a) secondary science, technology, engineering, and/or mathematics teachers or (b) elementary teachers who are specialists that teach only mathematics and/or science. The term *induction programs* refers to formal support programs provided by school districts or schools for 1st-year teachers and sometimes 2nd- and 3rd-year teachers. Such programs typically feature formal mentoring, orientations prior to the start of the school year, and regular workshops/professional development activities. Induction programs sometimes feature training for mentors, release time for 1st-year teachers, and teacher assessments. In contrast to *induction programs*, the term *induction* refers to the career stage that typically follows pre-service teacher preparation. This period usually lasts about two to four years or until the beginning teacher achieves more permanent employment status and/or a regular/professional teaching certificate or license. Individuals who enter teaching through alternative certification pathways and become full-time teachers of record while they are completing education coursework are often in the preparation and induction stages at the same time.

In the first section of this essay, we start by describing four large-scale, quantitative studies of how formal mentoring and induction programs affect beginning teacher retention (Glazerman et al., 2010; Kapadia, Coca, & Easton, 2007: Schmidt, Young, Cassidy, Wang, & Laguarda, 2017; Smith & Ingersoll, 2004). The next part of the first section considers quantitative and qualitative studies of how mentoring and induction is associated with changes in novice teachers' instructional quality. We conclude this first section by addressing how mentoring and induction influences beginning teacher effectiveness. The second and third sections examine how principal leadership and school organizational context are associated with teacher retention. See Figure 1 in the Appendix for a graphic representation of these three sections of the essay.

The fourth section identifies strengths and weaknesses of the methods and indicators used for research on mentoring/induction programs, principal leadership, and school organizational context and STEM teachers; and recommends new methods and indicators for research in this area. In the fifth section, we discuss challenges to research implementation in this area and we conclude with several suggestions for strengthening linkages among STEM teacher induction scholars. These include identifying models that will permit cross-comparison of findings across studies, using common research methods, and creating data collection guidelines.

We focused on the following main criteria in selecting studies to include in this review:

1. The data collected focus on beginning elementary and/or secondary teachers
2. The data were collected in the United States
3. The study was published between 2000 and 2018
4. For quantitative studies, the design included one or more of the following predictor variables: mentoring/induction program, principal leadership, teacher fit, social networks
5. For quantitative studies, the design included one or more of the following outcomes: beginning teacher retention, commitment, instructional quality, and/or effectiveness
6. For qualitative studies, the design addressed processes by which mentoring and induction programs seem to affect the experiences of beginning teachers

In this review, we included experimental studies and quantitative, non-experimental studies. While we searched for studies of teachers in STEM subjects, most of the included studies did not focus specifically on teachers who only teach STEM subjects.

Formal Mentoring and Induction Programs and STEM Teacher Retention, Instructional Quality, and Effectiveness

Researchers have used large-scale, quantitative data to examine the effects of different types of mentoring and induction programs on several outcomes including teacher retention, instructional practice, and effectiveness. They have also drawn on smaller-scale, qualitative data to investigate the processes by which and the conditions under which mentoring and induction programs can contribute to these outcomes. In this section, we first describe four large-scale, quantitative studies of elementary and/or secondary teachers across content areas (Glazerman et al., 2010; Kapadia, Coca, & Easton, 2007; Schmidt et al., 2017; Smith & Ingersoll, 2004) and review their results related to beginning teacher retention. Next, we consider the results of large-scale quantitative studies and small-scale qualitative studies of the effects of mentoring and induction on novice teachers' instructional practices (a) across content areas and (b) in STEM subjects. Finally, we review large-scale studies of the impact of mentoring and induction on beginning teacher effectiveness in mathematics and English language arts. Overall, we find some evidence that mentoring and induction programs affect beginning teacher retention, instruction, and effectiveness across content areas and in STEM subjects.

Mentoring, Induction, and Beginning STEM Teacher Retention

Smith and Ingersoll (2004) used data from the 1999-2000 Schools and Staffing Survey (SASS) and the Teacher Follow-Up Survey. The SASS is a nationally representative dataset; it is administered every three or four years and was administered in 1987-88, 1990-91, 1993-94, 1999-2000, 2003-04, 2007-08, and 2011-12. This survey is now known as the National Teacher and Principal Survey (NTPS); the NTPS was administered for the first time in 2015-16. A stratified probability sample design was used in the SASS to ensure sufficient numbers of reliable estimates in the samples of teachers, principals, schools, and districts. In their analyses, Smith and Ingersoll included 3,235 beginning elementary and secondary teachers from the 1999-2000 SASS; of these, 485 taught mathematics and/or science. These teachers were from public, charter, and private schools and their schools varied in terms of the percentages of low-income and racial/ethnic minority students that they served. The authors relied on observational data among first-year teachers in the SASS sample with regard to mentoring and other forms of induction support.

Smith and Ingersoll (2004) reported on whether first-year teachers had a mentor in the same field and whether they found their mentor to be helpful. But the SASS data did not include information about other characteristics of the mentors or how they were selected. In addition, the SASS data did not include details about the intensity, duration, cost, or structure of induction programs. Thus, in our review, we consider the beginning teachers in their sample to have received mentoring from full-time teachers who also taught in the same schools where the novices taught; as of 1999-2000, this was common in most U.S. school districts and it continues to be common in 2018-19.

Smith and Ingersoll (2004) reported that beginning middle school teachers were almost twice as likely as novice elementary teachers to leave after the first year while beginning high school teachers were about 50% more likely to leave. They also reported that novice mathematics and science teachers were about 10% more likely to leave teaching than other teachers, but this result was not statistically significant. The authors focused on first-year teacher retention. For their full sample of 3,235 beginning teachers, they found that having a mentor in one's field reduced the risk of leaving at the end of the first year by 30%, a result that was statistically significant at the 90% level of confidence. They did not indicate whether same-field mentoring had different effects on secondary STEM teachers compared to secondary non-STEM teachers, various types of secondary STEM teachers, elementary STEM teachers, or elementary non-STEM teachers. Also, the authors did not provide information about the percentages of different groups of teachers who were assigned out-of-field mentors (Smith & Ingersoll, 2004).

In contrast to Smith and Ingersoll (2004), Glazerman and colleagues (2010) conducted a randomized experiment that featured 418 elementary schools in 17 urban school districts. In their analyses, they included 1,009 beginning teachers in grades K-6 from urban, medium- to high-poverty school districts; of these teachers, 91% (i.e., more than 900) taught mathematics and were responsible for students' mathematics outcomes; the other 9% did not teach mathematics. Each of the 418 schools was randomly assigned to one of two treatment groups or a control group. In 10 of the 17 districts, the services (i.e., comprehensive induction) were offered to treatment schools for one year only. In the remaining 7 districts, the services were offered to treatment schools for two years. The authors compared beginning teachers who received comprehensive induction to beginning teachers who received low-intensity induction.

The two treatments included the ETS Pathwise induction model and the University of California Santa Cruz New Teacher Center (NTC) induction model. Both treatments featured full-time, trained mentors; formative teacher assessment; classroom observations; and release time. It is important to note that these mentors were typically assigned to a caseload of about 12 to 15 beginning teachers; for a given mentor, they were often assigned to work with novice teachers whose teaching assignment differed from the mentor's former teaching assignment with regard to content area and grade level; in other words, the Pathwise and NTC induction models featured generic mentoring as opposed to subject-specific mentoring. The control group featured mentoring from a full-time teacher; at the same time, beginning teachers and mentors in the control group were much less likely than those in the treatment group to receive release time and mentors were much less likely to receive mentor training, to conduct classroom observations of their mentees, or to use a formative teacher assessment in their work with them (Glazerman et al., 2010).

Glazerman et al. (2010) followed teachers through the end of their third year and collected data on their retention, literacy instruction, and effects on mathematics and reading achievement. In terms of effects of comprehensive induction, the authors presented results for the study participants' third year of teaching. They reported that comprehensive induction did not make beginning teachers feel more satisfied or prepared; and it did not make them more likely to stay in their schools, their districts, or the profession. Glazerman and colleagues supplemented their experimental analyses with regression analyses in which they grouped the teachers from the experimental and control groups together and made use of variation in the actual induction support that teachers across the study received. For these analyses, they employed four explanatory variables based on the number of years the teacher had an assigned mentor and measures of the instructional focus, intensity, and breadth of the mentoring that they received. None of the four measures of induction support were related to retention in the district or in the profession. At the same time, novices who taught the same grade as their mentor had lower rates of retention in the district and in the profession than those who did not have such matches; this finding is contrary to the authors' expectations (and our expectations) and merits further study.

In a third study, Schmidt, Young, Cassidy, Wang, and Laguarda (2017) conducted a randomized experiment that featured 629 first- and second-year elementary teachers and mathematics and ELA secondary teachers and 227 schools in Broward County Public Schools (BCPS), Florida and Chicago Public Schools (CPS). Of these teachers, 342 in 108 schools were randomly assigned at the school level to the treatment across the two districts and 287 in 119 schools were assigned to the control group. The treatment was the NTC induction model (described above). Similar to the Glazerman et al. study, the control group featured mentoring from a full-time teacher; novice teachers and mentors in the control group were much less likely than those in the treatment group to receive release time; and mentors were much less likely to receive mentor training, to conduct classroom observations of their mentees, or to use a formative teacher assessment in their work with them. Each beginning teacher received mentoring and induction support for their first two years of teaching. The authors reported no statistically significant differences between the treatment and control groups with regard to retention after the second year of teaching (Schmidt et al., 2017)

Other studies have examined how mentoring and induction programs affect beginning teacher commitment, which is a strong predictor of teachers' actual retention decisions. For example, Kapadia, Coca, and Easton collected survey data in 2004-05 from novice (i.e., first- and second-year) teachers in CPS. The authors found that novice elementary teachers (i.e., in grades K-8) receiving "strong" mentoring "were much more likely (than other novices) to report a good experience, intend to continue teaching, and plan to remain in the same school" (Kapadia, Coca, & Easton, 2007, p.28). *Strong* mentoring included assistance with teaching and assessment strategies, classroom management, CPS policies and procedures, observation and discussion of teaching, and communication with parents that beginning teachers found to be very helpful. Teachers receiving *average* levels of mentoring experienced most types of assistance and found them somewhat or very helpful and teachers receiving *weak* mentoring either received no mentoring assistance or they participated in some mentoring activities, but found them at best somewhat helpful (Kapadia, Coca, & Easton, 2007).

Mentoring, Induction, and Beginning Teacher Instructional Quality

While the prior part of this section addressed teacher retention, which is largely predicated on teachers' sense of satisfaction with their job, the next part focuses on how mentoring and induction programs are associated with the quality of beginning teachers' instructional practices. We first describe studies of novice teachers' instructional quality across content areas and then we focus on studies of novice STEM teachers' instructional quality.

Beginning Teachers' Instructional Quality Across Content Areas. As noted, Glazerman et al. (2010) observed each first-year teacher in their study as they taught a literacy lesson. The authors used the Diagnostic Classroom Observation (DCO) instrument to evaluate the lessons in three domains: implementation, content, and classroom culture. The first domain addressed such aspects of literacy implementation as pace, student choices, institutional choices, and best practices. The second domain focused on skill development, assessment, connections between reading and writing, and understanding content and close reading. The third domain addressed such aspects of literacy instruction as whether literacy was valued, clear and consistent routines, respectful behavior, and equal access to teacher and resources. The authors reported the psychometric properties of the DCO instrument, as employed in their study, and took steps to establish inter-rater reliability among the individuals who conducted the classroom observations (Glazerman et al., 2010).

Glazerman and colleagues (2010) observed no statistically significant differences between treatment and control teachers' performance in their first year on the three domains measured by the DCO instrument. They did not observe the teachers teach literacy lessons in their second or third years of teaching and they did not observe them teach mathematics lessons at all; this may explain why there were no significant effects of comprehensive induction on instructional quality.

In the study carried out by Schmidt and colleagues in BCPS and CPS, they used the Framework for Teaching (Danielson, 2013) to assess the classroom environments and instructional practices of the teachers in their study. In particular, they used this observation instrument to evaluate the beginning teachers' ability to promote student engagement, facilitate classroom discussion, communicate with students, assess student learning, establish a respectful and productive learning environment, and manage student behavior and classroom procedures.

The authors reported the psychometric properties of the Framework for Teaching, as used in their study. Schmidt and colleagues (2017) found no statistically significant differences between treatment and control teachers in their classroom practices. However, it's important to note that the Framework for Teaching does not use content-specific indices to generate sub-scores for teaching within subject domains. That is, this observation instrument can be used with any subject and it is employed the same way across different subjects.

Evertson and Smithey (2000) conducted a study of 46 mentor-first-year teacher pairs in two consortia of Midwestern school districts. One consortium featured nine school districts and 21 schools while the second consortium included 14 schools. There were 23 mentor-mentee pairs assigned to the treatment group and 23 pairs assigned to the control group; the authors did not indicate whether mentors and mentees were assigned to the two groups at random. The treatment group included seven teachers in grades Kindergarten through 4 while the control group featured eight teachers at this level. The treatment group included seven middle school teachers while the control group featured eight teachers at this level. And the treatment group included nine teachers at the high school level while the control group featured seven students at this level. The authors did not indicate whether any of the K-4 teachers taught mathematics and/or science only. In addition, the authors did not specify how many of the middle and high school teachers taught mathematics and/or science (Evertson & Smithey, 2000).

Mentors in both groups completed pre- and post-assessments of mentees' needs, weekly summaries of mentoring activities, and monthly reports of mentoring goals (Evertson & Smithey). The mentors in the treatment group also participated in a 4-day workshop on effective mentoring practices and monthly mentor meetings to address mentoring dilemmas and revisit mentoring concepts. The beginning teachers in the study were elementary, middle, and high school teachers, including some secondary mathematics and science teachers. Each novice teacher was observed three to six times each in the fall of their first year. The authors reported criterion-referenced reliability for individual observers in their study. Evertson and Smithey (2000) assessed their classroom climate, instructional management, use of rules and procedures, behavior management, the arrangement of their room, and student engagement. They reported that beginning teachers in the treatment group had better classroom organization and management practices than those in the control group, and their students were more engaged (Evertson & Smithey, 2000).

Stanulis and Floden (2009) collected data from 24 first-year teachers in a large school district, 12 of whom received intensive mentoring and 12 of whom received less intensive mentoring. Both groups included elementary teachers and secondary mathematics and science teachers. Beginning teachers in the intensive mentoring group met once a week with their mentors and attended monthly workshops with their mentors and other novice teachers; their mentors took part in mentor study groups for six hours each month as well as six days of professional development throughout the school year. The authors used the Atmosphere, Instruction/Content, Management, and Student Engagement (AIMS) classroom observation instrument, which generated ratings for each domain. The authors reported the psychometric properties of the AIMS instrument, as employed in their study, and took steps to establish inter-rater reliability among the individuals who conducted the classroom observations.

After eight months of mentoring, Stanulis and Floden (2009) found statistically significant differences between the first-year teachers in the intensive group and those in the second group with regard to atmosphere, instruction/content, and student engagement; in each case, the novices in the intensive mentoring group had higher scores than their counterparts in the other group. It is important to note that teachers were not randomly assigned to the two groups in this study.

Thompson, Goe, Paek, and Ponte (2004) studied the impact of the Beginning Teacher Support and Assessment (BTSA) program in California on beginning elementary teachers' English language arts (ELA) instructional practices. At the time of the study, all beginning teachers in California were required to receive support through local BTSA programs; the main components of BTSA were formal mentoring and the use of a formative teacher assessment. Local BTSA programs varied in the extent to which they included other induction components such as orientations prior to the school year, regular workshops, and release time. The authors surveyed third-year teachers in grades 3 through 5 from across California and they used the survey data to assign respondents to one of three categories: low, middle, or high levels of engagement with their local BTSA induction program. Subsequently, they observed subsets of the survey respondents as they taught two ELA lessons and interviewed them. The authors reported the psychometric properties of the Descriptions of Practice observation rubric, as used in their study, and took steps to establish inter-rater reliability among the individuals who conducted the classroom observations (Thompson, Paek, Goe, & Ponte, 2004).

Similar to Stanulis and Floden (2009), Thompson and colleagues (2004) reported that novice teachers who were highly engaged in induction had higher scores than the low engagement group on seven of nine measures of instructional practice. At the same time, the difference between high and low induction engagement teachers was statistically significant for only one of the measures (i.e., instructional planning). In addition, it's important to note that the participation rate in this study was only 26% (Thompson et al, 2004).

Beginning STEM Teachers' Instructional Quality. Luft, Roehrig, and colleagues have undertaken several studies of induction programs for beginning secondary science teachers. In one study, Luft, Roehrig, and Patterson (2003) collected data on three groups of secondary science teachers; each group included two first-year teachers, two second-year teachers, and two third-year teachers. One group participated in a science-oriented induction program, one took part in a general induction program, and one group did not receive formal induction support. Each teacher was interviewed eight times and observed teaching science three times over the course of one school year. Luft and colleagues (2003) reported that the study participants in the science-focused support program held beliefs aligned with student-centered inquiry science practices, enacted more student-centered inquiry lessons, and perceived fewer constraints in their teaching than did the other two groups of teachers.

In a subsequent study, Roehrig and Luft (2006) investigated the extent to which teacher preparation programs seem to influence the effects of a science-focused induction program on beginning secondary science teachers. The 16 study participants were all first-year secondary science teachers who had graduated the previous year from one of four different preparation programs. The researchers collected data on all of the participants during their first year of teaching – as they participated in the induction program – to examine their teaching beliefs, instructional practices, and experiences in the classroom. Roehrig and Luft found that the novice science teachers' pre-service training influenced the types of assistance they sought and took away from the induction program. In addition, during the year of data collection, those participants who had graduated from a preparation program with two science methods courses and an extended student teaching experience expressed beliefs in the efficacy of student-centered instruction and enacted more reform-based practices than the other teachers (Roehrig & Luft, 2006).

Luft et al. (2011) conducted longitudinal research into the effects of induction in the first and second years of teaching. They collected data from 98 secondary science teachers over a two-year period; 56 of the teachers took part in science-focused induction programs while 42 participated in general induction programs. The researchers conducted classroom observations of the study participants four times during each school year and each teacher took part in three types of interviews; one type of interview focused on their epistemological beliefs, one type addressed pedagogical content knowledge (PCK) in science, and one type occurred once a month and concentrated on the teachers' instructional practices. The authors took steps to establish inter-rater reliability among the individuals who conducted the classroom observations. Luft and colleagues (2011) reported that first-year teachers who participated in of the two science-focused induction programs strengthened their beliefs, PCK, and practices. While most of the teachers shared similar beliefs and PCK at the end of the second year, regardless of the induction program they participated in, the teachers in the science-focused induction programs continued to implement learning environments that were more interactive and had more laboratories and investigations than did the second-year teachers in the other induction programs (Luft et al., 2011).

It is important to note that researchers have grappled with the question of how best to measure the instructional quality of beginning STEM teachers. For example, Bieda, Salloum, Hu, Sweeny, Lane, and Torphy (under revision) compared and contrasted the relevant affordances and constraints of using the Instructional Quality Assessment (IQA; Matsumura et al., 2006) vs. Teaching for Robust Understanding of Mathematics (TRU Math; Schoenfeld, Floden, & the ATS Project Team, 2014) to assess the quality of novice elementary teachers' math instructional practices.

Other Research on Induction for Secondary STEM Teachers. Other studies have explored the processes by which mentoring and induction programs can affect the experiences of beginning secondary STEM teachers. Desimone and colleagues (2014) drew on survey and interview data to examine the experiences that 57 first-year middle school mathematics teachers had with formal and informal mentors. They reported that the beginning teachers (a) spent significantly more time interacting with their informal mentors and (b) were significantly more likely to focus in interactions with their informal mentors on expectations for teachers, parent involvement, and emotional support. Desimone and colleagues also found that the novice teachers spent more time interacting with formal mentors when these formal mentors had mathematics teaching experience. In addition, the authors reported that beginning teachers with more challenging teaching assignments spent more time interacting with their formal mentors in general and more often focused on mathematics instruction, classroom management, and being observed and receiving feedback. Finally, the study revealed that novices were more likely to spend time with formal and informal mentors who worked at their schools in contrast to mentors who were not working at their schools (Desimone et al., 2014).

In a different analysis from the same study, Polikoff, Desimone, Porter, and Hochberg (2015) used survey and interview data to investigate factors associated with high-quality mentoring interactions between beginning middle school mathematics teachers and their formal mentors. In a regression analysis, the authors reported that novice teachers who had time during the school day to meet with their formal mentors reported having higher-quality interactions with them compared to other novices (Polikoff et al., 2015). They also drew on interview data to explain that three aspects of school district mentoring policy seemed to be associated with novices' experiencing high-quality mentoring interactions: whether the mentor worked at their school and was located close to them within the building, whether they had time to meet with them, and whether the mentor had a formal evaluative role. In terms of mentors having evaluative roles, when mentors were responsible for contributing to formal evaluations of beginning teachers, this increased the likelihood that they observed them regularly and provided feedback to them on their instruction (Polikoff et al., 2015).

Mentoring, Induction, and Beginning Teacher Effectiveness

In the Glazerman et al. (2010) study, with regard to effects of comprehensive induction on teacher effectiveness, the authors presented results for the study participants' third year of teaching. For teachers in the 10 one-year districts, there was no impact of comprehensive induction on gains in their students' math or reading scores. For teachers in the seven two-year districts, the impacts on gains in math and reading scores were positive and statistically significant. In addition, in their supplemental (i.e., non-experimental) analyses, the authors found some positive, statistically significant associations between induction support and gains in student achievement in mathematics, but not for reading (Glazerman et al., 2010).

Schmidt, Young, Cassidy, Wang, and Laguarda (2017) examined the impact of two years of NTC induction support on the effectiveness of teachers in grades 4 through 8 in the areas of English language arts (ELA) and mathematics. In terms of ELA achievement gains, on average, students in grades 4 through 8 of teachers who received two years of induction support outperformed students of control teachers by 0.09 standard deviation ($p < .05$); this is comparable to moving from the 48th to the 52nd percentile. With regard to mathematics achievement gains, students in grades 4 through 8 of teachers who received two years of induction support performed 0.15 standard deviation ($p < .01$) higher on average than students of control teachers; this is similar to moving from the 46th to the 52nd percentile.

Summary

In synthesizing the results from these studies, several things become apparent. First, the findings from the two large-scale, randomized experiments do not provide evidence that comprehensive induction has a stronger effect on early career teacher retention than less intensive approaches to mentoring and induction (Glazerman et al., 2010; Schmidt et al., 2017). In contrast, results from some studies suggest that programs in which mentees are matched with school-based mentors who teacher full-time promote beginning teacher commitment and retention (Kapadia, Coca, & Easton, 2007; Smith & Ingersoll, 2004). Our interpretation is that it is generally helpful for first-year teachers to work with mentors who are full-time teachers at their schools and who have strong understandings of their teaching assignments (i.e., grade level, subject, and curricula) and of their school organizational contexts. On the other hand, working with mentors who have less understanding of their teaching assignments and school organizational contexts (as in the NTC and Pathwise models) does not have a noticeable positive impact on teacher retention.

Second, the findings from the two randomized experiments also do not provide evidence that comprehensive induction has a stronger effect on early career teachers' instructional practices than less intensive approaches to mentoring and induction (Glazerman et al., 2010; Schmidt et al., 2017). Results from smaller, non-experimental studies show some impact of comprehensive mentoring on instruction (Evertson & Smithey, 2000; Stanulis & Floden, 2009), but neither of these studies focused primarily on STEM instruction. Third, the randomized experiments by Glazerman et al. (2010) and Schmidt et al. (2017) indicate that two years of comprehensive induction has a statistically significant and positive impact on teacher effectiveness (as measured by student tests in mathematics and reading language arts). Our interpretation is that the first year of full-time teaching is extremely challenging for most teachers; as a result, robust interventions (e.g., comprehensive induction, well-designed pre-service teacher education, instructional coaching) are more likely to impact teacher effectiveness after two years (i.e., in the third year), if at all, than in the first two years of teaching. In other words, it can take two or more years to see the results of such interventions. This seems consistent with research on pre-service teacher preparation that has found effects of preparation experiences on second-year teachers' instructional practices and effectiveness (Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2009; Grossman et al., 2000).

Finally, it is important to note that few of the studies reviewed in this section focused solely on STEM teachers. With the exception of (a) research by Luft, Roehrig, and colleagues; and (b) Desimone, Polikoff, and colleagues; most prior research on mentoring and induction programs has concentrated on non-STEM teachers.

Principal Leadership and STEM Teacher Retention and Effectiveness

Several studies have drawn on large-scale data to examine the role of principal leadership in promoting the retention of STEM teachers, particularly early career STEM teachers. In contrast to research on mentoring and induction, these studies have consistently found that school leaders strongly influence this important outcome. At the same time, few studies have investigated ways in which principal leadership affects STEM teachers' instructional practice or effectiveness. In this section, we first review several large-scale studies of principal leadership and STEM teacher retention. Next, we describe a few studies that have explored associations between principal leadership and student achievement gains in mathematics.

Principal Leadership and Teacher Retention

Ingersoll and May (2012) used nationally representative data from the 2003-04 SASS and the 2004-05 Teacher Follow-up Survey to examine how the retention of qualified first-year secondary and elementary mathematics and science teachers was affected by various school factors. The authors defined *qualified* mathematics teachers as those who completed an undergraduate or graduate major in mathematics or mathematics education; and they defined *qualified* science teachers as those who completed an undergraduate or graduate major in science education, in an area of natural science, or in engineering. Ingersoll and May (2012) noted that 85% of the teachers in their study taught at the middle or high school levels. The remaining 15% worked at the elementary level; of these, the majority worked as single-subject teachers (i.e., teaching only mathematics or science) or as instructional coaches or specialists in mathematics or science.

The school factors in Ingersoll and May's study included administrative support, degree of student discipline problems, teacher influence over school decisions, teachers' classroom autonomy, the maximum salary offered by one's district, school resources, access to useful, content-focused professional development (PD), and access to useful PD related to student discipline. According to the authors, the strongest predictors of mathematics teacher retention "were the degree of individual classroom autonomy held by teachers; the provision of useful, content-focused PD; useful PD concerning student discipline and classroom management; and the degree of student discipline problems" (Ingersoll & May, 2010, p.39). The findings were similar for science teachers; the strongest predictors were useful, content-focused PD; the extent of student discipline problems; and the maximum salary offered by one's district. In sum, while teacher retention was not significantly affected by administrator support, other factors that principals can impact (i.e., provision of PD, degree of student discipline problems) were found to shape retention.

In research in New York City (NYC), Boyd, Grossman, Lankford, Loeb, and Wyckoff (2011) examined the impact of six aspects of school organizational context on the retention of all K-12 teachers who were in their first year of teaching in 2004-05. The teachers were surveyed at the end of their first and second years of teaching, and the authors also administered a follow-up survey to those individuals who participated in the study as first-year teachers and subsequently left teaching in NYC. Thus, the sample included teachers who returned to the same school (where they had taught in their first year) for their second year of teaching, teachers who moved to a different school in NYC for their second year, and individuals who left teaching in the district after their first year.

The study explored how teacher retention was affected by several factors, including school leadership, teachers' classroom autonomy and influence on school policies, teachers' relations with colleagues, student behavior, school safety, and school facilities. Boyd and colleagues (2011) found that school leadership was the only factor that significantly predicted teacher retention for both elementary and secondary teachers. Their measure of school leadership was based on several survey items including whether the school administration was viewed as supportive, whether an effective school discipline policy was in place, whether the school administration was perceived as evaluating teachers fairly, and whether the school administration consulted with faculty before making decisions.

In research in Chicago, Allensworth, Ponisciak, and Mazzeo (2009) investigated the effect of principal leadership and other school organizational conditions on the retention of all K-12 teachers who were in their first year of teaching in 2005-06. As in Boyd et al. (2011), the sample included teachers who returned to the same school for their second year of teaching, teachers who moved to a different school in Chicago for their second year, and individuals who left teaching in Chicago after their first year. The study found that in schools where teachers perceived the principal as a strong instructional leader, schools where teachers expressed high levels of trust in their principal, and schools where teachers reported having notable influence over school decisions, beginning teacher retention rates were higher (Allensworth, Ponisciak, & Mazzeo, 2009). These results applied to both elementary and secondary teachers. The authors added that when different school conditions were analyzed together, "some of the relationship between principal leadership and teacher stability (was) explained by other working conditions in the school, but that principal leadership (remained) a strong, significant predictor of teacher stability on its own" (Allensworth, Ponisciak, & Mazzeo, 2009, p.26). These other working conditions included collective responsibility, teacher influence, relational trust, and program coherence (Allensworth, Ponisciak, & Mazzeo, 2009).

In research in North Carolina, Ladd (2011) drew on a 2006 statewide survey administered to K-12 teachers at all levels of experience to explore the relationship between teachers' views of working conditions and their planned and actual departure rates from their schools. The author considered the effects of several working conditions on these outcomes: the quality of school leadership, expanded roles for teachers, PD for teachers, facilities and resources, and, at the elementary and middle school levels, time to meet with students and other teachers. Ladd's measure of school leadership quality was based on a number of survey items, including whether the principal was viewed as supportive with regard to instruction and student discipline, whether the principal maintained high expectations for student learning and teachers' instruction, whether teachers trusted the principal, whether the principal involved teachers in decision making, and, at the elementary and middle school levels, whether teachers viewed the teacher evaluation process as legitimate and fair. Ladd (2011) reported that the quality of school leadership was a stronger predictor of teachers' planned and actual departures than any other working conditions variable. Further, the quality of principal leadership had a stronger effect on teacher attrition than the percentage of racial/ethnic minority students or the percentage of students eligible for free or reduced price lunch.

Kim (2017) employed nationally representative data from the Beginning Teacher Longitudinal Study (BTLS) to examine how different types of principal leadership influenced early career teacher retention. The dataset for the BTLS featured all of the first-year teachers in the 2007-08 SASS, including elementary, middle, and high school teachers; and followed this set of teachers for five years, through 2011-12. These teachers included elementary teachers who taught one or more STEM subjects as well as secondary STEM teachers. The author first showed that principal leadership had a stronger impact on early career teacher retention in their schools, as compared to retention in the profession, and that this effect of school leadership remained constant during teachers' first five years in teaching. Second, Kim explored how different types of principal leadership predicted teacher retention in their schools, including instructional leadership, leadership related to managing student behavior, and leadership related to creating a supportive school climate. The author reported that principal leadership related to creating a supportive school climate had a significant negative association with teachers' likelihood of leaving the school. In contrast, though, with other studies reviewed here, Kim (2017) found that principal leadership related to instruction and behavior management did not have a strong association with novice teacher mobility.

Principal Leadership and STEM Teacher Effectiveness

Heck and Hallinger (2009) collected data from 195 elementary schools in a western state in the U.S. to examine how distributed leadership was associated with student achievement gains in mathematics. The teachers in their study included teachers at all experience levels; i.e., they did not limit their teacher sample to novice or early career teachers. Their measure of distributed leadership included three items that addressed the extent to which a given principal made collaborative decisions focused on educational improvement; emphasized school governance that promoted commitment, participation, and shared accountability; and involved teachers in evaluating the school's academic development. Heck and Hallinger (2009) reported that distributed leadership had indirect, positive effects on student achievement gains in mathematics.

Leithwood and Mascall (2008) drew on survey data from teachers in 90 elementary and secondary schools to investigate how collective leadership was associated with student achievement gains in mathematics and language arts. Similar to Heck and Hallinger (2009), the teachers in their study included teachers at all experience levels. The authors created a measure of collective leadership by asking teachers in a survey about the extent to which each of the following groups had direct influence on school decisions: teachers with designated leadership roles, other teachers, staff teams, students, principals, other school administrators, district administrators, parent advisory groups, and other parents. Thus, this notion of collective leadership is similar to the construct of distributed leadership referred to above. Also, it is important to note that principals play an important role in creating school climates in which teachers experience high levels of collective leadership. Leithwood and Mascall found that all school members and other stakeholders had higher levels of leadership influence in higher-achieving schools than they did in lower-achieving schools. These differences were most pronounced in terms of the leadership enacted by staff teams, students, and parents (Leithwood & Mascall, 2008).

Using district-wide data from 200 elementary schools in Chicago (serving grades K-8), Newmann, Smith, Allensworth, and Bryk (2001) examined how different aspects of principal leadership were associated with gains in students' mathematics and reading achievement. Newmann and colleagues reported that schools with high levels of instructional program coherence promoted higher gains in math and reading achievement than other schools that served students from similar demographic backgrounds. Schools with high levels of instructional program coherence were characterized by principals who (a) established rigorous student learning goals and coordinated curriculum and (b) provided professional development and carried out teacher evaluation in ways that reinforced the common instructional framework.

Grissom, Loeb, and Master (2013) drew on observation data from principals in about 125 elementary and secondary schools in Miami-Dade County to investigate how principals allocated their time across various instructional leadership activities and how different activities were associated with student achievement gains in mathematics and reading. The researchers reported that time devoted to (a) coaching teachers on instruction and (b) evaluating teachers or their curriculum was related to higher achievement gains. On the other hand, in contrast to time spent coaching or evaluating teachers, overall time spent by principals on general instructional leadership activities was not linked to school effectiveness. In addition, time devoted to brief classroom walkthroughs was not associated with achievement gains (Grissom et al., 2013).

Summary

We draw a number of conclusions from synthesizing the results of studies of the effects of principal leadership. For one, principal leadership has consistently been found to strongly influence beginning teacher retention; one can extrapolate this finding to argue that it likely applies to the retention of both secondary and elementary STEM teachers. Second, these studies indicate that several types of principal leadership are important: instructional leadership, including support for teacher PD; the ability to create a supportive school culture; leadership related to student discipline; and the ability and willingness to seek teacher input into decision making, establish and maintain trust with teachers, and evaluate teachers fairly (Allensworth, Ponisciak, & Mazzeo, 2009; Boyd et al., 2011; Ingersoll & May, 2012; Kim, 2017; Ladd, 2011). These types of leadership seem to be equally important for both secondary and elementary STEM teacher retention.

Finally, there is evidence that principals who enact distributed or collective leadership, who promote instructional program coherence, and/or who coach and evaluate teachers are likely to promote STEM teacher effectiveness among teachers at all experience levels; at the same time, it’s unclear whether or how distributed or collective leadership influence the effectiveness of beginning STEM teachers.

School Organizational Context, Commitment, Retention, and Instructional Quality

In addition to research on mentoring, induction, and principal leadership, a growing number of studies have examined how school organizational context affects beginning teacher outcomes. In this section, we review studies of person-environment fit, novices' social networks, and teacher commitment, retention, and instructional quality.

Recent studies have drawn on a long tradition in industrial-organizational (I-O) psychology of examining the degree of fit between teachers and their environments. For example, in research in Michigan and Indiana, Pogodzinski, Youngs, and Frank (2013) drew on data collected in 2007-08 from 144 early career elementary and middle school teachers in 11 school districts in the two states to assess the relationship between person-organization (P-O) fit and intent to continue teaching in one's school. In each participating district, all first-, second-, and third-year general education teachers in grades 1-8 were invited to participate. The authors measured perceived P-O fit with a series of six survey questions that addressed the degree of fit between the novice's approach to teaching, professional goals, and professional interests; and those of other teachers at their school. They found moderate correlations between teachers' perceived fit with their school and their commitment to their school.

Player, Youngs, Perrone, and Grogan (2017) used data from the 2011-12 SASS and the 2012-13 Teacher Follow-Up Survey to explore how first-year elementary and secondary teachers' perceptions of person-job (P-J) fit and principal leadership predicted their retention decisions. The authors measured perceived P-J fit with questions that addressed the extent to which the teacher, for example, felt enthusiastic about teaching, considered leaving teaching for another position, and felt working at their current school was rewarding. Player and colleagues measured several aspects of principal leadership including whether the principal was supportive, whether teachers were recognized for good work, whether the principal supported teachers with discipline issues, and whether they communicated their vision for the school to the staff. The authors reported that P-J fit and leadership predicted retention in the teaching profession and in one's school, respectively, for both elementary and secondary teachers (Player et al., 2017).

In a third study, Frank, Kim, Salloum, Bieda, and Youngs (under review) drew on longitudinal data on the math instructional practices and social networks of 137 early career elementary teachers (ECTs) to examine how the institutional forces of the Common Core and student testing pressures affect ECTs' math instruction. The authors found that the network members of ECTs played a critical role. When network members perceived the Common Core and testing pressures to support their own practices, ECTs' levels of ambitious mathematics instruction were more likely to increase. In contrast, when network members perceived the Common Core and testing pressures to be inhibiting, or when ECTs named no close colleagues or formally assigned mentors, ECTs' levels of ambitious mathematics instruction were more likely to decline. These effects were net of ECTs' engagement in professional development, mathematical knowledge for teaching, and levels of ambitious mathematics instruction at the beginning of the academic year. The results help us understand how the immediate networks of ECTs shape their responses to ubiquitous institutional pressures.

Summary

In looking across these studies, it is clear that early career teachers' perceptions of fit with their school organizations and with the teaching profession are associated with their levels of commitment to their school and retention in their profession (Player et al., 2017; Pogodzinski, Youngs, & Frank, 2013). In addition, emerging research indicates that novice elementary teachers' social networks can influence the quality of their mathematics instruction.

Strength and Weaknesses of Methods and Indicators Used in Research on Beginning Teacher Induction Programs

In this section, we describe strengths and weaknesses of the methods and indicators used in research on beginning teacher induction. We also identify implications of the research reviewed above for the types of measures and indicators that should be included in future studies and we discuss some new methods and indicators that researchers should consider using in future research on novice STEM teacher induction.

Strengths

The methods and indicators used in recent studies of beginning teacher induction have several strong points. First, some studies have employed nationally representative data from the Schools and Staffing Survey (SASS; now known as the National Teacher and Principal Survey) and the Beginning Teacher Longitudinal Study (BTLS). The use of nationally representative data enables researchers to generalize their findings to all beginning teachers in the U.S.; in addition, studies employing such data usually have very large sample sizes, which enable researchers to reduce errors in their analyses. Second, a few of these studies have featured experiments in which beginning teachers are randomly assigned to a treatment group (i.e., those who receive comprehensive induction) or a control group (i.e., those who receive less intensive mentoring); by randomly assigning novice teachers to treatment and control groups, scholars are able to greatly reduce the likelihood that other factors are affecting their outcomes of interest, thus enabling them to reliably make causal arguments.

Third, several studies have made use of rich, comprehensive data from states (such as North Carolina) or large urban school districts (such as New York City or Chicago); such datasets feature large samples of beginning teachers as well as data on several other factors. This enables researchers to control for the possible effects of these other factors (i.e., to reduce the likelihood that they are affecting their outcomes of interest). Fourth, a key part of the knowledge base on STEM induction programs comes from in-depth qualitative research studies. These studies have greatly increased our understanding of the conditions under which mentoring and induction lead to important outcomes such as beliefs about teaching, instructional practices, classroom management practices, and student engagement (Evertson & Smithey, 2000; Luft et al., 2011; Luft, Roehrig, & Patterson, 2003; Roehrig & Luft, 2006; Stanulis & Floden, 2009).

Weaknesses

Despite many advances in research on beginning teacher induction during the past 15 years, there are some areas where more research is needed. First, most studies have focused on multiple-subject elementary teachers as opposed to concentrating on (a) elementary teachers who only teach mathematics and/or science; and (b) secondary teachers in the areas of mathematics, science, engineering, and technology. As a result, there is little descriptive information about how many elementary teachers, for example, teach science, engineering, and/or technology. Second, there have been few large-scale studies of how mentoring and induction programs, principal leadership, or beginning teachers' social networks affect novice STEM teacher instructional quality. As discussed in the next section, it is very challenging to collect reliable, valid classroom observation data from large numbers of beginning STEM teachers. Nonetheless, more research is needed on how induction, leadership, and teacher fit with school organizational conditions can promote high-quality STEM instruction.

Third, there have been few studies of how mentoring and induction programs, principal leadership, person-environment fit, or social networks affect beginning STEM teacher effectiveness. In order to detect influences of induction on student learning, the studies by Glazerman et al. (2010) and Schmidt et al. (2017) suggest a need to follow novice teachers for at least two or three years. Fourth, there have been few studies of how beginning secondary STEM teacher fit with their schools or the teaching profession is associated with retention. Secondary mathematics and science teachers typically have a wide range of occupational opportunities available to them outside of K-12 teaching. Thus, in investigating their degree of fit with their schools and with teaching more generally, it would be important for researchers to examine reasons why novice STEM teachers entered teaching in the first place and how their teaching experiences compare to non-teaching work opportunities in terms of salary and working conditions.

Implications for Future Research

The findings from the studies reviewed above have several implications for future research on beginning STEM teacher induction programs. First, these study results indicate the importance of including measures of whether the formal mentor assigned to a novice STEM teacher (a) works at or is knowledgeable about the novice's school, (b) is knowledgeable about the novice's curriculum and teaching assignment, and (c) is readily accessible. Second, the results from these studies suggest that first-year teachers may have too many demands on their time to make productive use of mentoring and induction. Future research should include measures of first-year teachers' non-teaching responsibilities and the total amount of resources provided to them. It may be that second- and third-year teachers are in a much better position to make productive use of induction services than first-year teachers. Third, the results indicate that effective principal leadership for beginning STEM teacher induction is not always clearly visible to novice teachers or others. Thus, it's important for studies of induction to include measures of leadership related to creating a supportive school culture, student discipline, teacher evaluation, and school-based decision making. Finally, it is important for research on novice STEM teacher induction to include teachers in the areas of engineering, technology, and integrated STEM coursework.

New Methods and Indicators

There are a few promising methods and indicators that researchers should consider using in future research on beginning STEM teacher induction. First, several reliable, valid classroom observation instruments have been developed during the past 10 years; these include content-specific instruments, such as the Instructional Quality Assessment (IQA; Matsumura et al. (2006), the Mathematical Quality of Instruction (MQI; Hill, Kapitula, & Umland, 2011) instrument, the Teaching for Robust Understanding in Mathematics (TRU Math) instrument (Schoenfeld, Floden, & the Algebra Teaching Study and Mathematics Assessment Project, 2014), and the Mathematics-Scan (M-Scan; Berry et al., 2013). These instruments enable researchers to collect data on instructional quality at scale.

Second, social network analysis (SNA) enables researchers to better understand the resources available to beginning STEM teachers in their schools and how these resources are typically shared. Using SNA enables scholars to move beyond a focus on formal mentoring and induction programs to consider how informal sources (i.e., other teacher colleagues) can influence novice STEM teacher retention, instructional quality, and effectiveness.

Challenges to Research Implementation

While randomized experiments, classroom observation instruments, measures of teacher effectiveness, and nationally representative data all have great potential to help advance the knowledge base in this area, each type of research also faces challenges and has limitations. Similarly, both social network analysis and the experience sampling method hold great promise for research on novice STEM teacher induction, but both approaches are resource intensive.

In terms of collecting data directly from school districts, there are two main general challenges. The first involves securing permission from districts to conduct this type of research. This requires strong collaborative relationships between researchers and district- and/or state-level administrators to carry out these types of studies. Also, districts often have very legitimate reasons for not participating in such research studies. The second general challenge is that beginning teachers may agree to participate in a study initially, but then drop out of the study over time because they start teaching in a different school district, change their teaching assignment within the same district, leave teaching altogether, or choose to leave the study while remaining in teaching. In other words, researchers need to use effective strategies for retaining study participants over time while remaining cognizant that many novice teachers move to different positions early in their careers or leave teaching completely.

One challenge in conducting experimental studies of beginning STEM teacher induction is attaining a sufficient number of novice secondary teachers in mathematics, science, engineering, and/or technology in order to carry out statistical analyses. One of the two large-scale experimental studies described above (Glazerman et al., 2010) focused on multiple-subject elementary teachers (and did not include elementary mathematics or science specialist teachers or secondary STEM teachers) while the other (Schmidt et al., 2017) featured multiple-subject elementary and secondary mathematics and ELA teachers, but for most analyses did not disaggregate the results for secondary teachers based on subject area taught. A second challenge is the need to conduct longitudinal research over two or more years in order to detect effects of STEM induction. Both of these studies (Glazerman et al., 2010; Schmidt et al., 2017) reported no effects of comprehensive induction after one year of teaching, but some impact on teacher effectiveness after the second or third year of teaching.

A main challenge in using classroom observation instruments to measure beginning teacher instructional quality in large-scale studies is the need to train multiple raters to score lessons reliably, to make sure they have sufficient knowledge of subject matter and subject matter pedagogy, and to calibrate them over time to maintain inter-rater reliability. A related challenge involves whether to use a generic observation tool or a subject-specific observation tool. Generic instruments would be advantageous in research on novices' instruction in multiple subjects (e.g., high school mathematics, biology, chemistry, and physics; or elementary mathematics and science) while subject-specific tools would be more useful in single-subject studies of instruction (e.g., elementary or secondary mathematics or science).

There are two main challenges in studying beginning STEM teacher induction and teacher effectiveness. First, while it is possible in virtually all states to collect student mathematics achievement data for teachers in grades 3-8, it is less feasible to collect such data for teachers in early elementary mathematics (grades K-2), elementary school science, middle school science, or high school mathematics or science. Second, as noted, Schmidt et al. (2017) reported effects of comprehensive induction on student learning after two years of teaching while Glazerman et al. (2010) found such effects after three years of teaching; thus, studies of induction and teacher effectiveness would need to be carried out over at least two or three years.

A primary challenge in using nationally representative data to study novice STEM teacher induction is that datasets such as SASS generally do not include refined measures of mentoring, induction, or principal leadership. For example, in working with the 1999-2000 SASS, Smith and Ingersoll (2004) noted that this dataset included little information about the intensity, duration, cost, or structure of the mentoring and induction programs experienced by the teachers in their study.

There are two main approaches to collecting social network data for beginning STEM teachers. Collecting egocentric network data involves (a) asking focal novice STEM teachers to identify teachers and other colleagues who have been important resources for them in teaching STEM and then (b) collecting data on these educators' knowledge, instructional practices, and/or expectations. In contrast, collecting sociometric network data involves administering surveys to all of the teachers in a focal novice STEM teachers' department or school and using the department- or school-wide to examine sources of STEM resources and expertise, how resources and expertise are shared, and whether focal STEM teachers have access to them. In both approaches, researchers use data on social network members to examine how their knowledge, instructional practices, or other characteristics affect beginning STEM teacher outcomes. At the same time, both approaches are resource-intensive. For example, in large-scale studies of 150 or more beginning STEM teachers, it can be prohibitively expensive to collect sociometric data for each novice teacher, especially if they are teaching at 100 or more different schools.

Strengthening Linkages Among STEM Induction Researchers

In this concluding section, we present a series of recommendations designed to strengthen linkages among STEM induction researchers. First, we describe next steps for research on mentoring and induction programs as well as instructional coaching. Second, we discuss future research on principal leadership, person-environment fit, and social networks. Third, we identify a number of common research methods and data collection guidelines that scholars should consider using or following in future research.

Research on Mentoring and Induction Programs

Based on our review of the literature, we identified two main approaches to formal mentoring and induction programs. The first approach, comprehensive induction, involves training formal mentors, providing them with release time from teaching, assigning them to work with as many as 10 to 15 beginning teachers, having them observe novices' instructional practices, and having them use formative teacher assessments to provide feedback to their mentees. In contrast, less intensive mentoring and induction typically features mentoring from a full-time teacher in the mentee's school; it is less common for such approaches to include mentor training, release time, classroom observations, or formative assessments.

Research by Glazerman et al. (2010) and Schmidt et al. (2017) indicates that comprehensive induction in mid-size to large, high-poverty urban districts has little impact on beginning teacher retention or instructional quality; after two or three years of teaching, comprehensive induction influences teacher effectiveness in mathematics at the elementary and middle school level. At the same time, there has been less research on the effects of comprehensive induction (a) in suburban and rural school districts or (b) on high school mathematics or science teachers, middle school science teachers, or engineering or technology teachers. Also, despite numerous small-scale studies of induction and beginning STEM teacher instructional quality, there has been relatively little large-scale research on this topic.

Thus, one direction for future research would be for scholars to study on a large scale mentoring and induction programs in suburban and rural districts. Another direction would be to study such programs for high school mathematics and science teachers, middle school science teachers, engineering teachers, and technology teachers. A growing number of school districts are offering courses in engineering and technology as well as integrated STEM courses. And a third direction would be to use content-specific observation instruments to study how comprehensive and less intensive mentoring programs are associated with changes in instructional quality during novice teachers' first three years of full-time teaching.

Research on Instructional Coaching

A growing number of studies have found that instructional coaching has strong effects on teachers' instructional practices and student learning, especially in STEM subjects (e.g., Allen, Hafen, Gregory, Mikami, & Pianta, 2015; Blazar & Kraft, 2015; Campbell & Malkus, 2011; Kraft, Blazar, & Hogan, 2018). Therefore, a key question for future research is why coaching tends to have such an impact on changes in instruction and teacher effectiveness in STEM subjects, especially among early career teachers, while mentoring and induction programs seem to have less impact on such changes.

Research on Principal Leadership and School Organizational Context

Our review of the literature also indicated that several types of principal leadership have a strong influence on beginning STEM teacher retention. This includes leadership related to instruction, creating a supportive school culture, classroom management, teacher evaluation, and school-based decision making. There have been few studies, though, of principal leadership and beginning STEM teacher instructional quality and relatively few studies of leadership and novice STEM teacher effectiveness.

Further, our review of the literature suggests that beginning STEM teachers' perceptions of fit with their schools and with the teaching profession are associated with their commitment and retention decisions. At the same time, there has been little research on how perceptions of person-environment fit or teachers' social networks are associated with beginning STEM teaching instructional quality or effectiveness.

Common Research Methods and Data Collection Guidelines

Large-scale quantitative research on beginning STEM teacher induction typically features surveys of novices; it also sometimes includes classroom observation data, student achievement data, and/or surveys of other educators. In contrast, small-scale qualitative research can involve interviews, classroom observations, surveys, and logs of mentoring activities.

In future research on novice STEM teacher induction, researchers should use classroom observation instruments that have evidence of reliability and validity and they should train and calibrate raters to ensure inter-rater reliability. In addition, when feasible, we recommend that researchers employ subject-specific observation tools such as the MQI or TRU Math. Such observation instruments focus on content-specific teaching knowledge and skills that teachers need to enact ambitious instructional practices and help students learn rigorous content and develop conceptual understanding. For example, the MQI focuses on the richness of mathematics lessons, working with students and mathematics, and mathematical errors and imprecisions (Blazar, 2015). In collecting student achievement data, we recommend that researchers collect such data over two or more years and account for missing data.

Finally, we believe that social network data (collected through surveys or interviews) is a promising research method in this area. In collecting social network data, we recommend that researchers use similar instruments with novice teachers and their social network members and take steps to ensure high rates of survey or interview completion.

References

Allen, J. P., Hafen, C. A., Gregory, A. C., Mikami, A. Y., & Pianta, R. (2015). Enhancing secondary school instruction and student achievement: replication and extension of the My Teaching Partner-Secondary intervention. *Journal of Research on Educational Effectiveness*, *8,* 475–489.

Allensworth, E., Ponisciak, S., & Mazzeo, C. (2009). The schools teachers leave: Teacher mobility in Chicago Public Schools. Chicago: Consortium on Chicago School Research.

Atteberry, A., Loeb, S., & Wyckoff, J. (2017). Teacher churning: Reassignment rates and implications for student achievement. *Educational Evaluation and Policy Analysis, 39*(1), 3-30.

Berry, III, R.Q., Rimm-Kaufman, S.E., Ottmar, E.M., Walkowiak, T.A., Merritt, E., & Pinter, H.H. (2013). The Mathematics Scan (M-Scan): A measure of mathematics instructional quality. Charlottesville, VA: University of Virginia.

Bieda, K., Salloum, S., Hu, A., Sweeny, S., Lane, J., & Torphy, K. (under revision). Capturing early career elementary teachers' enactment of mathematics practice at scale.

Blazar, D. (2015). Effective teaching in elementary mathematics: Identifying classroom practices that support student achievement. *Economics of Education Review, 48,* 16-29.

Blazar, D., & Kraft, M. A. (2015). Exploring mechanisms of effective teacher coaching: A tale of two cohorts from a randomized experiment. *Educational Evaluation and Policy Analysis*, *37*(4), 542-566.

Boyd, D., Grossman, P., Ing, M., Lankford, H., Loeb, S., & Wyckoff, J. (2011). The influence of school administrators on teacher retention decisions. *American Educational Research Journal, 48*(2), 303-333.

Boyd, D.J., Grossman, P.L., Lankford, H., Loeb, S., & Wyckoff, J. (2009). Teacher preparation and student achievement. *Educational Evaluation and Policy Analysis, 31*(4) (pp.416-420 and pp.428-436 required).

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

Campbell, P. F., & Malkus, N. N. (2011). The impact of elementary mathematics coaches on student achievement. *The Elementary School Journal*, *111*(3), 430-454.

Danielson, C. (2013). *The framework for teaching evaluation instrument.* Princeton, NJ: The Danielson Group.

Desimone, L.M., Hochberg, E.D., Porter, A.C., Polikoff, M.S., Schwartz, R., & Johnson, L.J. (2014). Formal and informal mentoring: Complementary, compensatory, or consistent? *Journal of Teacher Education, 65*(2), 88-110.

Evertson, C.M., & Smithey, M.W. (2000). Mentoring effects on proteges' classroom practice: An experimental field study. *Journal of Educational Research, 93*(5), 294-304.

Frank, K., Kim, J., Salloum, S., Bieda, K., & Youngs, P. (under review). How beginning elementary teachers draw on networks to navigate competing institutional forces.

Glazerman, S., Isenberg, E., Dolfin, S., Bleeker, M., Johnson, A., Grider, M., Jacobus, M. & Ali, M. (2010). *Impacts of Comprehensive Teacher Induction: Final Results from a Randomized Controlled Study*. Washington, DC: U.S. Department of Education, Institute of Education Sciences.

Grissom, J.A., Loeb, S., & Master, B. (2013). Effective instructional time use for school leaders: Longitudinal evidence from observations of principals. *Educational Researcher, 42*(8), 433-444.

Grossman, P.L., Valencia, S.W., Evans, K., Thompson, C., Martin, S., & Place, S. (2000). Transitions into teaching: Learning to teach writing in teacher education and beyond. *Journal of Literacy Research, 32*(4), 631-662.

Harris, D., & Sass, T. (2011). Teacher training, teacher quality, and student achievement. *Journal of Public Economics, 95*, 798–812.

Heck, R. H., & Hallinger, P. (2009). Assessing the contribution of distributed leadership to school improvement and growth in math achievement. *American Educational Research Journal*, *46*(3), 659-689.

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

Ingersoll, R.M., & May, H. (2012). The magnitudes, destinations, and determinants of mathematics and science teacher turnover. *Educational Evaluation and Policy Analysis, 34*(4), 435-464.

Johnson, S.M., 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.

Kapadia, K., Coca, V., & Easton, J.Q. (2007). Keeping new teachers: A first look at the influences of induction in the Chicago Public Schools. Chicago: Consortium on Chicago School Research, University of Chicago.

Kim, J. (2017). Quality matters: The influence of teacher evaluation policies and school context on teaching quality. Unpublished doctoral dissertation, Michigan State University, East Lansing, MI.

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.

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.

Lankford, H., Loeb, S., & Wyckoff, J. (2002). Teacher sorting and the plight of urban schools: A descriptive analysis. *Educational Evaluation and Policy Analysis*, *24*(1), 37-62.

Leithwood, K., & Mascall, B. (2008). Collective leadership effects on student achievement. *Educational Administration Quarterly*, *44*(4), 529-561.

Luft, J.A., Firestone, J.B., Wong, S.S., Ortega, I., Adams, K., & Bang, E. (2011). Beginning secondary science teacher induction: A two-year mixed methods study. *Journal of Research in Science Teaching*, *48*(10), 1199-1224.

Luft, J.A., Roehrig, G.H., & Patterson, N.C. (2003). Contrasting landscapes: A comparison of the impact of different induction programs on beginning secondary science teachers' practices, beliefs, and experiences. *Journal of Research in Science Teaching*, *40*(1), 77-97.

Matsumura, L.C., Slater, S.C., Junker, B., Peterson, M., Boston, M., Steele, M., & Resnick, L. (2006). Measuring reading comprehension and mathematics instruction in urban middle schools: A pilot study of the Instructional Quality Assessment. CSE Technical Report 681. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

Newmann, F.M., Smith, B., Allensworth, E., & Bryk, A.S. (2001). Instructional program coherence: What it is and why it should guide school improvement policy. *Educational Evaluation and Policy Analysis*, *23*(4), 297-321.

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.

Player, D., Youngs, P., Perrone, F., & Grogan, E. (2017). How fit is associated with teacher mobility and attrition. *Teaching and Teacher Education, 67,* 330-339.

Polikoff, M. S., Desimone, L. M., Porter, A. C., & Hochberg, E. D. (2015). Mentor policy and the quality of mentoring. *The Elementary School Journal*, *116*(1), 76-102.

Pogodzinski, B., Youngs, P., & Frank, K. (2013). Collegial climate and novice teachers' intent to remain teaching. *American Journal of Education*, *120*(1), 27-54.

Roehrig, G.H., & Luft, J.A. (2006). Does one size fit all? The induction experience of beginning science teachers from different teacher preparation programs. *Journal of Research in Science Teaching*, *43*(9), 963-985.

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

Scafidi, B., Sjoquist, D. L., & Stinebrickner, T. R. (2007). Race, poverty, and teacher mobility. *Economics of Education Review*, *26*(2), 145-159.

Schmidt, R., Young, V., Cassidy, L., Wang, H., & Laguarda, K. (2017). Impact of the New Teacher Center's new teacher induction model on teachers and students. Menlo Park, CA: SRI international.

Schoenfeld, A. H., Floden, R. E., & the Algebra Teaching Study and Mathematics Assessment Project. (2014). *The TRU Math scoring rubric*. Berkeley, CA & E. Lansing, MI: Graduate School of Education, University of California, Berkeley & College of Education, Michigan State University.

Smith, T.M., & Ingersoll, R.M. (2004). Reducing teacher turnover: What are the components of effective induction? *American Educational Research Journal*, *41*(3), 681-714.

Stanulis, R.N., & Floden, R.E. (2009). Intensive mentoring as a way to help beginning teachers develop balanced instruction. *Journal of Teacher Education*, *60*(2), 112-122.

Thompson, M., Goe, L., Paek, P., & Ponte, E. (2004). *Study of the impact of the California Formative Assessment and Support System for Teachers: Report 1, Beginning teachers' engagement with BTSA/CFASST.* (CFASST Rep. No. 1, ETS RR-04-30). Princeton, NJ: Educational Testing Service.

Thompson, M., Paek, P., Goe, L., & Ponte, E. (2004). *Study of the impact of the California Formative Assessment and Support System for Teachers: Report 2, Relationship of BTSA/CFASST engagement and teacher practices.* (CFASST Rep. No. 2, ETS RR-04-31). Princeton, NJ: Educational Testing Service.

Appendix

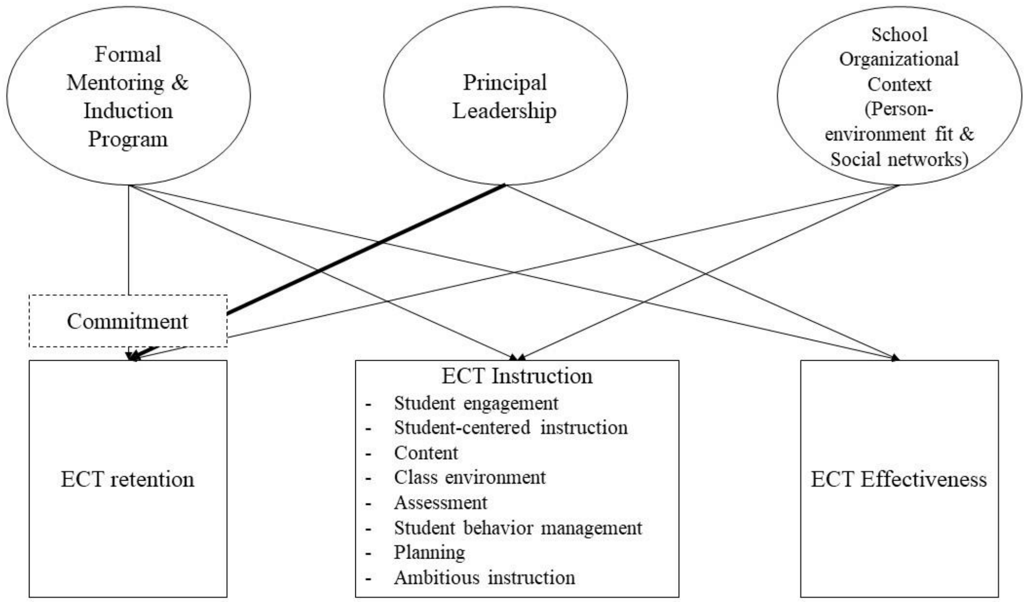

Figure 1. Teacher Induction Programs and Other Factors that Lead to Teacher Retention in the STEM Teaching Workforce

This figure summarizes the main sections of research reviewed in this essay. The essay examines how formal mentoring and induction programs are associated with (a) early career teacher (ECT) retention; (b) ECTs' instructional practices; and (c) ECTs' effectiveness. In addition, the essay considers how principal leadership is associated with (a) and (c). Finally, the essay explores how school organizational context is associated with (a) and (b).