Review of Educational Research
http://rer.aera.net

The Impact of Induction and Mentoring Programs for Beginning Teachers: A Critical Review of the Research

Richard M. Ingersoll and Michael Strong

REVIEW OF EDUCATIONAL RESEARCH published online 18 April 2011
DOI: 10.3102/0034654311403323

The online version of this article can be found at:
http://rer.sagepub.com/content/early/2011/04/15/0034654311403323

Published on behalf of

American Educational Research Association

and

http://www.sagepublications.com

Review of Educational Research
Season XXXX, Vol. XX, No. X, pp. xx–xx
DOI: 10.3102/0034654311403323

The Impact of Induction and Mentoring Programs for Beginning Teachers: A Critical Review of the Research

Richard M. Ingersoll
University of Pennsylvania

Michael Strong
University of California, Santa Cruz

This review critically examines 15 empirical studies, conducted since the mid-1980s, on the effects of support, guidance, and orientation programs—collectively known as induction—for beginning teachers. Most of the studies reviewed provide empirical support for the claim that support and assistance for beginning teachers have a positive impact on three sets of outcomes: teacher commitment and retention, teacher classroom instructional practices, and student achievement. Of the studies on commitment and retention, most showed that beginning teachers who participated in induction showed positive impacts. For classroom instructional practices, the majority of studies reviewed showed that beginning teachers who participated in some kind of induction performed better at various aspects of teaching, such as keeping students on task, using effective student questioning practices, adjusting classroom activities to meet students' interests, maintaining a positive classroom atmosphere, and demonstrating successful classroom management. For student achievement, almost all of the studies showed that students of beginning teachers who participated in induction had higher scores, or gains, on academic achievement tests. There were, however, exceptions to this overall pattern—in particular a large randomized controlled trial of induction in a sample of large, urban, low-income schools—which found some significant positive effects on student achievement but no effects on either teacher retention or teachers' classroom practices. The review closes by attempting to reconcile these contradictory findings and by identifying gaps in the research base and relevant questions that have not been addressed and warrant further research.

KEYWORDS: beginning teacher induction, mentoring programs, teacher mentors.

For decades, education researchers and reformers have called attention to the challenges encountered by newcomers to school teaching. However, traditionally teaching has not had the kind of support, guidance, and orientation programs for new employees—collectively known as induction—common to many skilled

blue- and white-collar occupations and characteristic of the traditional professions (Lortie, 1975; Tyack, 1974; Waller, 1932). Although elementary and secondary teaching involves intensive interaction with youngsters, the work of teachers is done largely in isolation from colleagues. School reformers and researchers have long pointed out that this isolation can be especially difficult for new teachers, who, on accepting a position in a school, are often left on their own to succeed or fail within the confines of their own classrooms—often likened to a "lost at sea" or "sink or swim" experience (e.g., Johnson, 1990; Johnson & Birkeland, 2003). Others go further, arguing that newcomers often end up placed in the most challenging and difficult classroom and school assignments—akin to a "trial by fire" experience (e.g., Lortie, 1975; Sizer, 1992). Indeed, some have assailed teaching as an occupation that "cannibalizes its young" (Ingall, 2006, p. 140).

Perhaps not surprisingly, teaching has also traditionally been characterized as an occupation with high levels of attrition among newcomers (Lortie, 1975; Tyack, 1974). All organizations and occupations, of course, experience some loss of new entrants—either voluntarily because newcomers decide to not remain or involuntarily because employers deem them to be unsuitable. Moreover, some degree of employee turnover, job, and career change is normal and inevitable.

However, teaching has relatively high turnover compared to many other occupations and professions, such as lawyers, engineers, architects, professors, pharmacists, and nurses (Ingersoll, 2003; Ingersoll & Perda, 2011), and teacher turnover is especially high in the first years on the job. Several studies have calculated that between 40% and 50% of new teachers leave within the first 5 years of entry into teaching (e.g., Grissmer & Kirby, 1987, 1992, 1997; Hafner & Owings, 1991; Ingersoll, 2003; Murnane, Singer, Willett, Kemple, & Olsen, 1991).

Recent research has also documented that one of the negative consequences of these high levels of turnover in teaching is their link to the teacher shortages that seem to plague schools perennially. In analyses of national data, we have found that neither the much heralded mathematics and science shortage nor the minority teacher shortage is primarily the result of an insufficient production of new teachers, as is widely believed. In contrast, the data indicate that these school staffing problems are to a significant extent a result of a "revolving door"—where large numbers of teachers depart teaching long before retirement (Ingersoll & May, 2011; Ingersoll & Perda, 2010; also see Achinstein, Ogawa, Sexton, & Freitas, 2010). Moreover, the data show that beginning teachers, in particular, report that one of the main factors behind their decisions to depart is a lack of adequate support from the school administration.

These are the kinds of occupational ills that effective employee orientation and induction programs seek to address, and in recent decades a growing number of states, school districts, and schools have developed and implemented induction support programs for beginning teachers. Our background analyses of national data show that the percentage of beginning teachers who report that they participated in some kind of induction program in their first year of teaching has steadily increased over the past two decades—from about 40% in 1990 to almost 80% by 2008. By 2008, 22 states were funding induction programs for new teachers (Education Week, 2008).

The theory behind induction holds that teaching is complex work, that preemployment teacher preparation is rarely sufficient to provide all of the knowledge

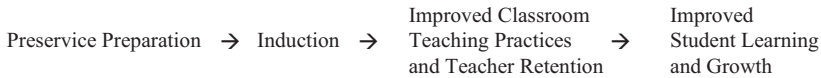

FIGURE 1. *Theory of teacher development.*

and skill necessary to successful teaching, and that a significant portion can be acquired only while on the job (see, e.g., Feiman-Nemser, 2001; Ganser, 2002; Gold, 1999; Hegstad, 1999). Hence, this perspective continues, there is a necessary role for schools in providing an environment where novices are able to learn the craft and survive and succeed as teachers. The goal of these support programs is to improve the performance and retention of beginning teachers, that is, to both enhance and prevent the loss of teachers' human capital, with the ultimate aim of improving the growth and learning of students (see Figure 1).

Typical of theory underlying induction is Zey's (1984) mutual benefits model, drawn from social exchange theory. This model is based on the premise that individuals enter into and remain part of relationships to meet certain needs, for as long as the parties continue to benefit. Zey extended this model by adding that the organization as a whole (in this case the school) that contains the mentor and mentee also benefits from the interaction.

From this theoretical perspective, teacher induction is distinct from both preservice and in-service teacher professional development programs. Preservice refers to the education and preparation candidates receive before employment (including clinical training, such as student teaching). In-service development refers to periodic upgrading and additional professional development received on the job, during employment. Theoretically, induction is intended for those who have already completed basic preemployment education and preparation. These programs are often conceived as a "bridge" from student of teaching to teacher of students. Of course, these theoretical distinctions can easily become blurred in real situations.

Although the overall goal of these teacher development programs is to improve the performance and retention of beginning teachers, parallel to the induction processes common to other occupations, induction theorists have identified multiple objectives and emphases such programs may hold (e.g., Feiman-Nemser, 2001; Ganser, 2002). Among them are teacher socialization, adjustment, development, and assessment. For instance, some programs are primarily developmental and designed to foster growth on the part of newcomers; in contrast, others are also designed to assess, and perhaps weed out, those deemed ill suited to the job. Moreover, teacher induction can refer to a variety of different types of activities for new teachers—orientation sessions, faculty collaborative periods, meetings with supervisors, developmental workshops, extra classroom assistance, reduced workloads, and, especially, mentoring. Mentoring is the personal guidance provided, usually by seasoned veterans, to beginning teachers in schools. In recent decades, teacher mentoring programs have become a dominant form of teacher induction (Britton, Paine, Raizen, & Pimm, 2003; Fideler & Haselkorn, 1999; Hobson, Ashby, Malderez, & Tomlinson, 2009; Strong, 2009); indeed, the two terms are often used interchangeably.

The overall objective of teacher mentoring programs is to give newcomers a local guide, but the character and content of these programs also vary widely. Duration and intensity, for example, may be very different from program to program. Mentoring programs can vary from a single meeting between mentor and

mentee at the beginning of a school year to a highly structured program involving frequent meetings over a couple of years between mentors and mentees who are both provided with release time from their normal teaching loads. Programs also vary according to the number of new teachers they serve; some include anyone new to a particular school, even those with previous teaching experience, whereas others focus solely on novices. Finally, mentoring programs vary as to how they select, prepare, assign, and compensate the mentors themselves. How carefully mentors are selected is an issue for programs, as is whether selection to be a mentor is truly voluntary or a semimandatory assignment. Some programs include training for mentors; some programs do not. Programs differ according to whether and how they pay mentors for their services. Some programs devote attention to the match between mentor and mentee; others do not. For instance, some programs may strive to see that new secondary-level math teachers are provided with mentors who have had experience teaching secondary-level math.

What kinds of induction and mentoring programs exist and under which circumstances they help are fundamental questions for researchers, educators in the field implementing such programs, and policymakers faced with decisions about supporting such programs. For the latter groups especially, investing in beginning teachers poses a conundrum. On one hand, as induction theory holds, investments that enhance the effectiveness of new teachers can add to the attractiveness of the job, improve teacher retention, and improve other outcomes. On other hand, if a significant portion of those entering teaching view it as a temporary line of work and plan to leave soon, regardless of such enhancements, the investments in human capital could be lost to the school.

These issues and concerns have gained increased attention in recent years—perhaps partly because of downturns in the larger economy and a greater emphasis on accountability and partly because of changes in the character of the teaching force itself. After two decades of flat growth, since the mid-1980s the teaching force in the United States has dramatically increased in size. This upsurge in hiring has resulted in an equally dramatic growth in the number of newly hired, first-year teachers the past two decades—from 50,000 in 1987–1988 to 200,000 in 2007–2008. In the late 1980s the modal teacher had 15 years of teaching experience; by 2008, the modal teacher was a beginner in his or her first year of teaching. Moreover, those data show that the attrition rates of first-year teachers—now the largest group within the occupation—have slightly increased over the past two decades (Ingersoll & Merrill, 2010). In short, both the number and instability of beginning teachers have been increasing.

For all of these reasons, with the growth of induction and mentoring programs, there has also been mounting interest in empirical research on the variety and value of these initiatives. Over the past couple of decades, numerous studies have been done on different types of programs. However, it is unclear how much of this research warrants unambiguous conclusions about the value of the induction program being considered. Some studies appear to lack methodological rigor and draw conclusions that reach beyond what their data truly support. Moreover, the content, duration, and delivery of programs vary so much from one site to another that it is not clear to what extent general conclusions about induction can be drawn from the research. Hence, there is a need to critically assess the empirical research on teacher induction to determine its scope and merit and the conclusions that may be drawn from it.

A number of useful reviews on the topic of induction have been published over the past two decades (for a recent anthology, see Wang, Odell, & Clift, 2010). Many of these reviews have focused on the theory, rationale, and conceptualization of induction (e.g., Feiman-Nemser, 2001; Feiman-Nemser & Schwille, 1999; Ganser, 2002; Gold, 1999; Hegstad, 1999). Others have focused primarily on the character of specific teacher induction reforms and initiatives (e.g., Fideler & Haselkorn, 1999; Scherer, 1999; Serpell & Bozeman, 1999; Wang & Odell, 2002). Still others examined teachers' experiences with induction (e.g., Wang, Odell, & Schwille, 2008). At least one review studied the conditions that give rise to effective mentoring and looked at the benefits of mentoring for both mentors and mentees (Hobson et al., 2009). However, there have been few efforts to provide comprehensive and critical reviews of empirical studies that evaluate the effects of induction on various outcomes. In 2004, we released an online review of empirical research on mentoring, in particular, and its effects on one outcome—teacher retention (Ingersoll & Kralik, 2004). In 2009, a second critical assessment of induction research appeared (Strong, 2009). The present review updates and expands these two earlier efforts by including more recent research and by broadening the purview to include studies on the effects of induction in general and on outcomes beyond teacher retention alone. Our objective is to provide researchers, policymakers, and educators with a reliable and current assessment of what is known and not known about the effectiveness of teacher induction and mentoring programs. Our objective is also to identify gaps in the research base and pinpoint relevant questions that have not been addressed and that warrant further research.

Review Method

We began by contacting leading researchers in the field and analysts in state governmental agencies. We examined existing systematic, narrative, or traditional reviews of such research, and we searched online databases including Dissertation Abstracts, ERIC, Psychological Abstracts, Sociological Abstracts, PsycINFO, Wilson Index, SAGE's online database, and Google Scholar. In the online searches we used combinations of three key terms—*beginning teacher induction*, *mentoring programs*, and *teacher mentors*—with several other terms—*program evaluation*, *teacher improvement*, *effectiveness*, *retention*, *student achievement*, and *teaching practice*. In our search, we included both published and unpublished documents on teacher induction and studies both from the United States and from other countries. Interest in teacher induction and mentoring appeared to gain momentum in the mid-1980s; hence, our review focuses on studies from that period to the present.

Our initial search located more than 500 documents concerned with teacher induction and mentoring. These included essays, reviews, monographs, reports, and articles. In a second step, we excluded all documents that were not empirical studies reporting data on beginning teacher induction and mentoring programs—trimming our list to about 150 documents. We then took a closer look at the documents themselves and excluded any of these empirical studies that failed to meet any of three criteria. This step resulted in a further reduction to 15 studies selected for this review (see Table 1). To the best of our knowledge, these 15 studies, forming the core of this review, exhaust the evidence base concerning the effects of teacher induction, insofar as the evidence meets the following criteria.

(Text continues on p. 10.)

TABLE 1
Fifteen Studies of ohe Effects of Induction (Marked With Asterisks in the References Section)

Overview	Data	Outcomes	Findings
Effects of induction on beginning teacher commitment and retention			
1. Kapadia et al. (2007) Evaluated districtwide induction programs in Chicago Public Schools for 2005; looked at data from 1,737 novice teachers (72% of all 1- to 2-year teachers); identified weak, average, and strong intensity of induction	Teacher questionnaires	How positive was first year; intentions to stay in teaching and/or in same school	Strong induction showed significantly higher scores in all three outcomes; no induction showed no difference from induction
2. Fuller (2003); Cohen and Fuller (2006) Evaluation for 1999–2003 of TxBESS, a statewide program to provide support for beginning teachers, of which mentoring was a major component	Annual questionnaire to mentees; state database on teacher retention	Teacher retention compared with other teachers in the state	TxBESS teachers were retained at significantly higher rates over first 3 years compared to other teachers in the state
3. Henke et al. (2000) Secondary analysis of Baccalaureate and Beyond Survey that followed a nationally representative sample of 7,294 college graduates who entered teaching after 1992–1993 school year; follow-ups in 1994 and 1997	Survey that included one yes–no question about induction	Attrition	Significantly lower attrition (15% vs. 26%) for beginning teachers who participated in induction program

(continued)

TABLE 1 (continued)

Overview	Data	Outcomes	Findings
4. Smith and Ingersoll (2004); Ingersoll and Smith (2004a, 2004b) Secondary analysis of nationally representative sample from the Schools and Staffing Survey and Teacher Follow-up Survey of 3,235 first-year teachers in 1999–2000 school year	Mailed questionnaire	Attrition after first year	Significantly lower attrition for teachers having different types of induction supports such as a helpful mentor in the same subject area or participation in collaborative activities with other teachers; no decrease in attrition for teachers receiving a reduced teaching load or a teacher aide in the first year
5. Hahs-Vaughn and Scherff (2008) Secondary analysis of subsample of English teachers from the 1999–2000 Schools and Staffing Survey	Mailed questionnaire	Individual/school characteristics on attrition, mobility, and retention	No effects for induction, but authors failed to limit analysis to first-year teachers; therefore, results are problematic
6. Duke et al. (2006) Secondary analysis of 1999–2000 Schools and Staffing Survey	Mailed questionnaire	Amount of induction; teacher intentions to stay	Induction had positive effect on teacher intentions to stay, but authors failed to limit analysis to first-year teachers; therefore, results are problematic
Effects of teacher induction on beginning teachers' classroom practices			
7. Evertson and Smithey (2000) Compared the effects of having trained versus untrained mentors; randomly assigned 46 teachers to each group	Classroom observations	Classroom practice	Teachers with trained mentors had better classroom organization and management early in the year, and students were more engaged
8. Roehrig et al. (2008) Case studies of six novice teachers and their mentors	Surveys, observations using AIMS instrument and interviews	Classroom practice	Ambiguous findings; both more and less effective teachers declined in use of effective practices over the year

(continued)

TABLE 1 (continued)

Overview	Data	Outcomes	Findings
9. Davis and Higdon (2008)			
Two groups of five teachers were studied; one group had a university-supplied mentor as well as district support; the other group had district support only	Two half-day observations in fall and spring; survey looked at mentor support	Classroom practice	School–university induction partnerships "may" contribute to teacher effectiveness
10. Stanulis and Floden (2009)			
Two matched groups of 12 beginning teachers; treatment group had intensive mentoring supplied by university; comparison group had district only support	Classroom observation early and late in year using AIMS instrument	Classroom practice	Experimental group showed gains in AIMS scores over year that were greater than the comparison group
Effects of teacher induction on student achievement			
11. Thompson et al. (2004)			
Studied California BTSA program among 1,125 third- to fifth-grade teachers from 107 school districts during their third teaching year	Survey of all teachers, interviews and observations of subsample	Engagement in BTSA and teaching practice; student achievement	Found high engagement in BTSA was associated with higher scores on most measures of teaching practice; students of teachers with higher engagement had higher test scores
12. Fletcher et al. (2008)			
Compared beginning teachers from three California school districts with different levels of BTSA induction support	Student test data; school district data; induction program data	Student achievement gains	Found teachers in the most intensive induction program had greater gains in reading; also, teachers in the intensive program showed class gains equal to those of experienced teachers in the same district

(continued)

TABLE 1 (continued)

Overview	Data	Outcomes	Findings
13. Fletcher and Strong (2009)			
Compared two groups of beginning teachers in an urban school district: those with full-time mentors and those with part-time mentors; all mentors had the same training	Student test data; district data; induction program data	Student achievement gains	Teachers supported by full-time mentors showed greater achievement gains over 1 year than those with part-time mentors
14. Rockoff (2008)			
Studied the effects of a comprehensive mentoring program provided by the Santa Cruz New Teacher Center on teachers in New York City in 2004	Survey and other data from the mentoring program; payroll data; NY DOE survey; standardized test data	In-school retention; teacher self-report on effectiveness; student achievement	Retention a function of previous experience in that school; teachers claimed mentoring affected teaching; more time with mentor showed higher achievement in math and reading
15. Mathematica Policy Research (four reports): Glazerman et al. (2006); Glazerman et al. (2008); Isenberg et al. (2009); Glazerman et al. (2010)			
Randomized controlled study of comprehensive induction support (adapted from two prominent induction programs) versus standard district support; recruited 17 large school districts with at least 50% low-income students; initial sample of 1,009 teachers; subsample followed for a second year; some analysis after 3 years	Observation, interview, questionnaire, and student test data; outside agency monitored treatment implementation	Intensity of induction support; teacher retention; teacher practice; student achievement	Treatment group received significantly more intensive induction support; no effects on retention, practice, or student achievement after 1 year; no effects on retention or achievement after 2 years; student achievement of treatment teachers significantly higher after 3 years (for small subsample)

Note: TxBESS = Texas Beginning Educator Support System; AIMS = The Atmosphere, Instruction/Content, Management, and Student Engagement measure; BTSA = Beginning Teacher Support and Assessment program; NY DOE = New York City Department of Education.

Evaluation and Outcomes

We included only empirical studies that sought to evaluate the effects of induction using one or more outcomes. We excluded empirical studies that were descriptive rather than evaluative, that is, studies that sought solely to summarize or describe the extent, process, content, or character of induction programs (e.g., Fideler & Haselkorn, 1999; Ganser, 1994, 1996; Schaffer, Stringfield, & Wolfe, 1992; Wollman-Bonilla, 1997). This meant that we excluded research on induction that focused solely on the "lived experiences" of teachers (Hobson et al., 2009). We recognize that firsthand accounts from beginning teachers on the content and processes of induction programs may provide rich information, but we elected to concentrate on studies that provided evidence of effects. We also excluded evaluative studies that focused on outcomes other than the effects of induction programs on beginning teachers or their students. For example, we excluded research that examined the factors, policies, and conditions that affect the provision of quality induction (e.g., Youngs, 2007) and omitted studies that evaluated only the effects of mentoring programs on mentors themselves.

Comparisons

We included only evaluative studies of induction that compared outcome data from both participants and nonparticipants in particular induction components, activities, or programs. The majority of empirical studies we initially examined were reports of program evaluations that collected data on outcomes solely from those who had participated in the induction programs being evaluated (e.g., Gregson & Piper, 1993; Mitchell & Scott, 1998; Stroot et al., 1999; Villeme, Hall, Burley, & Brockmeier, 1992; Wilson, Darling-Hammond, & Berry, 2001). Such studies can provide valuable feedback to providers of, and participants in, such programs, but they cannot offer unambiguous conclusions about the effects of participating or not participating.

Some studies selected for our review were able to compare those participating in induction to those who did not participate in induction. However, since induction has become widespread, most of the studies we review compare teachers according to their *degree* of participation, that is, those with more or less participation in one or more induction components, activities, or programs. To use a medical research analogy, most the studies reviewed here are not the equivalent of research that compares taking aspirin to not taking aspirin but of research that compares taking different dosages of aspirin or taking aspirin versus taking other drugs.

Explicit Description of Data and Method

We included only studies that contained explicit descriptions of their data sources, sample sizes, research methods, and outcomes. For instance, we excluded studies whose outcomes were not sufficiently well defined or measured for us to assess the accuracy of the results (e.g., Bradley & Gordon, 1994; Perez, Swain, & Hartsough, 1997). In the case of quantitative studies, we also included only those providing tests of statistical significance, where possible and appropriate.

Studies Reviewed

The studies we review vary in their data and method. Some were evaluations of specific district or state mentoring programs. Some involved close-up examination

of small samples of classrooms. Others used secondary analysis of large-scale databases to statistically investigate the association of induction with outcomes. The nature of the data reported across the studies reviewed did not permit a meta-analysis without eliminating a significant number of studies, along with the useful information they provide.

Compared to some other topics, such as school size (e.g., Leithwood & Jantzi, 2009), the evidence base for this review is relatively small. Given the diminished sample size, we are able to summarize a selection of the studies in some detail, elaborating the strengths and limitations of each. The outcomes of the studies we review fall into three broad categories: (a) teachers' job satisfaction, commitment, retention, and turnover, (b) teachers' classroom teaching practices and pedagogical methods, and (c) student achievement. Our review is organized in three sections, corresponding to these three major sets of outcomes. The exception is the largest study to date—a randomized controlled trial that investigated the impact of comprehensive induction on all three sets of outcomes (Glazerman et al., 2010)—which we review in a separate section.

The Effects of Induction on Beginning Teacher Commitment and Retention

In this section we focus on studies (see Table 1) that provide evidence about the relationship between participation in induction and a beginning teacher's job satisfaction, commitment, retention, or turnover. Three studies were evaluations of specific state or school district beginning teacher induction programs. Four involved secondary statistical analyses of large-scale nationally representative teacher surveys.

In most of the studies, the investigators examined data on teachers' actual retention or departures obtained from surveys of individual teachers, districts, or state personnel databases. In two studies, the investigators used as an outcome beginning teachers' self-reported *intentions* regarding how long they planned to remain in teaching rather than teachers' actual retention or turnover. It is unclear how closely self-reported intentions mirror actual retention behavior; this measure most likely captures teachers' degree of commitment and job satisfaction rather than their longevity per se.

Evaluations of State and District Mentoring Programs

All three evaluations of specific school district or state beginning teacher induction programs found that induction had positive effects. That is, beginning teachers who received some type of induction had higher job satisfaction, commitment, or retention. We describe the two most thorough of these studies in some detail below.

In 2005, Kapadia, Coca, and Easton (2007) evaluated districtwide induction programs in Chicago Public Schools. They analyzed data for 1,737 novice teachers, representing 72% of the first- and second-year teachers employed in the district in 2005. The researchers divided the levels of induction and mentoring support that each teacher received into three groups: weak, average, and strong. Interestingly, even though induction was compulsory in the school district, about one fifth of the teachers reported that they were not involved in any induction program. The researchers measured the influence of participation in induction programs on three self-reported teacher outcomes: how positive was a teacher's first year on the job, teachers' intentions to stay in teaching, and their intentions to stay

in the same school. The study used multilevel logistics regression for its analysis and was able to control for the background characteristics of teachers, classrooms, and schools, including working conditions that could affect the outcomes. Comparing those who received some level of induction to the 20% who reported receiving none, the study found that participation in induction, by itself, had little effect on any of the three outcomes. However, among those who received some level of induction, teachers in the strong induction group showed higher levels on all three outcomes. Mentoring was an important component, especially at the elementary level, but comprehensive induction, comprising multiple supports, had the most effect on intentions to remain in the same school. Kapadia et al. concluded that programs should focus on selection and training of mentors to ensure high levels of support and that teacher collaboration and principal assistance are the most influential factors for novices.

A second study evaluated the Texas Beginning Educator Support System (TxBESS; Cohen & Fuller, 2006; Fuller, 2003; also see Charles A. Dana Center, 2001). Begun in 1999, TxBESS was a statewide comprehensive program of instructional support, mentoring, and formative assessment to assist teachers during their first years of service in Texas public schools. School districts had discretion in selecting participants for the program. About 15% of the state's new teachers were involved. A key program objective was to improve retention of beginning teachers. The study obtained information from TxBESS participants through an annual mailed survey questionnaire. Among other things, the survey sought information on the nature of the relationship between mentors and mentees, including time spent with mentor, whether release time was granted (to both mentor and mentee) for these meetings, whether the mentee wanted a mentor, and the nature of the meetings with the mentor (e.g., formal vs. ad hoc, provided assistance with classroom management, assisted with learning the "unwritten rules" of the school, etc.). The study obtained data on teacher retention from a state personnel database and compared annual retention rates of TxBESS participants to those of all beginning teachers in the state from 1999–2000 through 2002–2003.

Analysis showed that among teachers who entered in the 1999–2000 school year, TxBESS participants left the Texas public school system at statistically significantly lower rates, for each of their first 3 years, than did teachers who did not participate in TxBESS. On disaggregating the data, the researchers found that these effects held up (in both magnitude and statistical significance) in both high-poverty and high-minority-enrollment schools. This was an important finding because these schools more often used the state program and had disproportionate numbers of beginning teachers in the TxBESS program but also generally had higher attrition of new teachers. Moreover, the analysis found that the retention effects held up across school levels; elementary, middle, and high schools all had significantly higher retention of TxBESS participants. Finally, the analysts also found that TxBESS appeared to help underqualified beginning teachers. TxBESS participation by beginning teachers who did not hold full certification, or who had been assigned to teach subjects out of their certification, resulted in better retention than when similarly underqualified teachers did not participate in TxBESS.

The TxBESS study has several limitations worth noting. First, since school districts selected participants for the program in different ways, differences in the characteristics

of participants and nonparticipants, rather than the program itself, might account for differences in outcomes. Second, since school districts differed in which components they used, variations in program content could account for different outcomes. Third, this study did not control for other factors that could also affect teacher retention, regardless of the existence of an induction or mentoring program.

Secondary Analyses of Large-Scale Nationally Representative Data

In addition to evaluations of specific induction programs, we also reviewed four studies that undertook secondary analyses of large-scale, nationally representative databases from the National Center for Education Statistics of the U.S. Department of Education to investigate the statistical association between induction and teacher retention. Three of the four studies found positive effects of induction; beginning teachers who received some type of induction had higher commitment to continuing as teachers or had higher retention. One study found no effects, but as we discuss below, this analysis, along with one of the studies showing positive effects, had serious flaws that undermined its validity.

In 2000, the National Center for Educational Statistics published an analysis undertaken by Henke, Chen, and Geis that used the 1993 Baccalaureate and Beyond Survey (B&B:93) to examine the experiences of new teachers, including the relationship between beginning teachers' participation in induction programs and their attrition. The B&B is a longitudinal survey that followed a nationally representative sample of those who graduated from undergraduate institutions in the 1992–1993 academic year. This cohort was interviewed during their senior year in 1993, interviewed in 1994 for a first follow-up, and interviewed a third time in 1997 for a second follow-up. The base sample who participated in all three interviews comprised 7,294 students. Henke et al.'s analysis focused on the experiences of the 7,294 college graduates from the class of 1992–1993 who entered elementary or secondary teaching.

Of the teachers in this sample, 46% reported participating in a school induction program when they entered teaching. The analysis revealed that about one fifth of recent college graduates who entered teaching between 1993 and 1997 were no longer teaching by July 1997; it also showed that participation in induction was negatively related to attrition from the occupation, at a statistically significant level. Of those who had participated in induction, 85% had stayed in teaching, compared to 74% of those who had not participated.

The B&B findings provide evidence from a nationally representative survey that teacher induction is related to lower teacher attrition. However, there are several important limitations to the B&B data and to the Henke et al. (2000) analysis. First, the item on teacher induction was a simple yes–no question and provided no detail on the type, characteristics, and components of induction. There is, for example, no way of knowing whether the induction program included a mentoring component. Second, the B&B survey focused on teachers fresh out of college with no prior teaching experience. This group is a subset of all those hired into teaching jobs that year and, hence, only a portion of those who did or did not participate in induction programs in any given year. Third, the Henke et al. analysis of the relationship between induction and attrition is based on bivariate correlations of one factor with the other and does not control for, or hold constant, other factors that

could account for differences in teacher attrition and for any apparent connection between teacher induction and teacher attrition.

A second study used data from the 1999–2000 Schools and Staffing Survey (SASS) and its supplement, the 2000–2001 Teacher Follow-up Survey (TFS), to analyze the relationship between participation in various induction activities and the retention of beginning teachers (Ingersoll & Smith, 2004a; Smith & Ingersoll, 2004). SASS is a nationally representative survey of teachers and administrators from public and private schools. The same schools were again contacted 12 months after the administration of the original SASS questionnaires, and all those in the original teacher sample who had moved from or left their teaching jobs were given a second questionnaire to obtain information on their departures. This latter group, along with a representative sample of those who stayed in their teaching jobs, constituted the TFS. The 2000–2001 TFS sample comprised about 7,000 elementary and secondary teachers; the study focused solely on beginning teachers—those without prior experience and in their first year of teaching in 1999–2000—a national sample of 3,235.

The analysis examined the association of three sets of induction-related measures drawn from an extensive battery of such items in the teacher survey questionnaire. The first set asked teachers whether they were working closely with a master or mentor teacher and, if so, whether the mentor was in the same subject area. The second set asked teachers whether they had any of the following collective supports: (a) seminars or classes for beginning teachers, (b) regular or supportive communication with their principal, other administrators, or their department chair, (c) common planning time or regularly scheduled collaboration with other teachers on issues of instruction, and (d) participation in a network of teachers (e.g., one organized by an outside agency or over the Internet). The third set of items asked teachers whether they received additional help to support their transition, including (a) a reduced teaching schedule, (b) a reduced number of preparations, and (c) extra classroom assistance (e.g., teacher aides).

The study's primary question was, does receiving any of these supports improve teacher retention? To answer this question, the researchers undertook a series of multinomial logistic regression analyses of the association between receiving these supports and the likelihood of beginning teachers' moving or leaving at the end of their first year on the job. To rule out other factors that might account for the observed effects of induction, the models included controls for numerous characteristics of teachers and their schools. After controlling for these background characteristics, the authors found that induction support was significantly associated with teachers' likelihood of turnover. But the analysis also found that the strength of the association depended on the type and number of supports. The strongest factors were having a mentor from the same field, having common planning time with other teachers in the same subject, and having regularly scheduled collaboration with other teachers. The weakest factors were a reduced teaching schedule, a reduced number of preparations, and extra classroom assistance.

The data also revealed that induction supports, activities, and practices rarely exist in isolation. In other words, of beginning teachers who had some kind of induction, most received several types of support. To look at the collective impact of receiving more than one support, the researchers tested the effects of packages or bundles of

supports on retention. The components selected for each package were based on how many teachers received them and the strength of their association with retention. The results showed that, collectively, as the number of components in the packages increased, the probability of turnover decreased, but also that the number of teachers receiving the package decreased. Participation in these activities, collectively, had a very large impact—the probability of a departure at the end of their first year for those getting a comprehensive package was less than half that of those who participated in no induction activities.

This analysis offers strong findings, especially for the advantages of bundles and packages of multiple induction components. One advantage of large-scale teacher databases, such as the SASS/TFS, is that they allow national assessments of whether a number of components of induction are associated with teachers' moving and leaving, after controlling for key background characteristics of teachers and their schools. However, there are important limitations to the 1999–2001 SASS/TFS database and to this study.

First, the questionnaire items provide limited depth and detail on the content and character of teacher induction and mentoring. For example, the survey asked teachers which kinds of supports their schools provided, but little information was obtained on the intensity, duration, cost, or structure of induction programs—information of vital importance to policymakers who must choose among many models. The analysis tells us, for example, that beginning teachers with mentors from the same field were less likely to leave after their first year, but many very different kinds of programs were no doubt lumped together in the responses to the mentoring question. It is likely that some of these programs were highly effective, some were moderately effective, and others were not effective at all. The analysis was not able to discern among them. Similarly, although the 1999–2000 SASS asked teacher mentees to evaluate how helpful their mentors were, little else was obtained on the characteristics of the mentors. Some observers have argued that the mere presence of a mentor is not enough; the mentors' knowledge of how to support new teachers and skill at providing guidance are also crucial (e.g., Evertson & Smithey, 2000; Kyle, Moore, & Sanders, 1999). These are important policy issues that the SASS data cannot address.

Second, although the statistical models in this study controlled for a wide range of teacher and school factors, the study did not control for or rule out other organizational and working conditions that likely exist in schools with higher quality induction packages and also affect turnover.

In a subsequent unpublished follow-up to this study, Ingersoll and Smith (2004b) disaggregated the 1999–2001 SASS/TFS to examine levels and effects of induction by school poverty levels. They found that the amount of induction received and its effect on turnover varied by the schools' poverty level. Their data revealed that teachers in high-poverty schools were at least as likely as, if not more likely than, their counterparts in low-poverty schools to receive and participate in induction and mentoring. The effect of these activities on reducing turnover, however, differed by school poverty level. Although the likelihood of leaving teaching at the end of the first year was significantly less in low-poverty schools where new teachers were matched with a mentor and had opportunities to collaborate with other teachers, the impact of these activities on retention in high-poverty schools

was small and statistically insignificant. Furthermore, although participation in a combined comprehensive package, or a greater number of induction activities, was associated with higher retention in low-poverty schools, this was not the case in high-poverty schools. The investigators concluded that either the quality of these programs differed substantially between high- and low-poverty schools or the organizational context in which new teachers enter teaching differed so dramatically between low- and high-poverty schools that the latter require different approaches to the socialization and support of new teachers. Unlike the earlier analysis, this second follow-up study controlled for a wide range of other organizational and working conditions, such as the quality of school leadership, the degree of student discipline problems, and the amount of faculty input into decision making. Positive levels of these factors were likely to coexist in schools with higher quality induction packages and to affect turnover. Interestingly, however, controlling for these factors did not change the initial findings—that induction had strong effects in low-poverty schools but not in high-poverty schools.

We reviewed two other studies that also analyzed data from the 1999–2000 SASS to examine the relationship between induction and retention. However, both studies had serious flaws in their data samples and analytic methods, making their findings of limited usefulness. The 1999–2000 SASS limited the questionnaire items on induction to teacher–respondents in their 1st through 5th years of teaching, as of the year of the survey. The first study, by Hahs-Vaughn and Scherff (2008), further restricted their analytic subsample to English teachers who, during the 1999–2000 school year, were in their 1st through 4th years of teaching, that is, the four cohorts who began teaching between the 1996–1997 and 1999–2000 school years, yielding a small sample of 86. The objective of their analysis was to assess the relationship between the amount of induction these four cohorts of beginning teachers experienced during their first year and the likelihood they would move or leave in later years. They found that induction had little effect.

SASS is a cross-sectional survey, and the TFS is only a 1-year longitudinal survey—it resurveys the original SASS sample 12 months later. The 1999–2001 SASS/TFS collected data from a sample of all those teaching in 1999–2000 and whether they moved or left between the 1999–2000 and 2000–2001 school years. Hence, for cohorts who entered before the 1999–2000 school year, the 1999–2001 SASS/TFS includes only *those still teaching as of 1999–2000*; by definition, it excludes those in earlier cohorts who moved or left in prior years. Hence, the study cannot assess the impact of induction on turnover of cohorts of teachers in their 1st through 4th years of teaching because those in their 2nd through 4th years who had already departed were no longer in the sample. In other words, the SASS/TFS data do not support longitudinal analysis of more than one cohort, as Hahs-Vaughn and Scherff (2008) sought to do. Analyses using the SASS/TFS to examine the effect of an intervention such as induction on turnover must necessarily focus on first-year teachers.

A similar problem holds for Duke, Karson, and Wheeler (2006). They, too, used the 1999–2000 SASS, and their objective was also to assess the impact of induction (along with field of undergraduate degree) on beginning teachers. Rather than actual turnover, they used as their outcome teachers' reports of how long they intended to remain in teaching. They found that induction had a positive impact on

teachers' plans to stay. Although their subsample was larger than Hahs-Vaughn and Scherff's (2008), Duke et al. also failed to limit their analysis to first-year teachers. Their analysis, like that of Hahs-Vaughn and Scherff, suffers from the same data censoring problem, thus also making the findings of limited usefulness.

The Effects of Teacher Induction on Beginning Teachers' Classroom Practices

We also reviewed studies (see Table 1) that provide evidence about the relationship between participation in induction and how well beginning teachers teach, including their skill, practices, development, and pedagogical methods. The strength of these studies is their close observation of teachers' actual behavior in classrooms or their careful assessment of teachers' practices through some kind of reflective interview. However, such data collection can be time-consuming, and the studies here necessarily focused on small teacher samples (from 6 to 287 teachers). A limitation of small samples, of course, is their low generalizability, and two of the five studies did not include tests of statistical significance (Davis & Higdon, 2008; Roehrig, Bohn, Turner, & Pressley, 2008). Studies that attempt to measure teachers' practices also face serious issues of validity and reliability and can encounter cognitive issues related to the observation of human behavior (for a discussion, see Strong, 2009).

None of these studies compared teachers who participated in induction to teachers who did not participate. In each of the five studies, all teachers in the sample participated in some induction, but the amount varied. Hence, the analyses compared teachers according to the degree and type of support they received from the program in their district. Four studies focused on the effects on beginning teachers of having different types of mentors. One of these four examined the effects of having trained mentors compared to having untrained mentors (Evertson & Smithey, 2000); two of the four examined the effects of receiving the existing district or school-based mentoring compared to having an additional mentor supplied by the study (Davis & Higdon, 2008; Roehrig et al., 2008); the fourth study examined the effects of receiving the existing district induction program (entailing mentoring, orientation, and seminars) compared to receiving intensive mentoring provided through a school–university partnership (Stanulis & Floden, 2009).

All of these studies used a variety of classroom teacher observation instruments that focused on aspects of classroom atmosphere, instructional methods, and classroom management. They all undertook at least two, and often three, classroom observations of each teacher, usually lasting several hours. Only one of the four studies randomly assigned participants to treatment and control groups (Evertson & Smithey, 2000). With one exception, all of the studies reported positive effects for their induction and mentoring treatment group. The exception (Roehrig et al., 2008) had ambiguous findings; beginning teachers regardless of induction intensity declined in their use of effective teaching practices over the course of their first year, but the more intensive group had a smaller decline than that the less intensive group.

The largest and most ambitious of this group of five studies (Thompson, Paek, Goe, & Ponte, 2004) is worth describing in some detail since it is unique in both

approach and sample size. In 2002, this research team was commissioned to study the impact of California's Beginning Teacher Support and Assessment program (BTSA) and its accompanying California Formative Assessment and Support System for Teachers (CFASST). All new teachers in California are required to receive BTSA support. Mentoring is the core element of this program, along with formative assessment. Other components of induction are optional, and BTSA programs vary widely across the state. Thus, the study compared teachers according to how much support they actually received. The study focused on the impact of the program on the teaching practices of beginning teachers and on the learning of their students.

The study surveyed the entire population of 1,125 third- to fifth-grade public school teachers in the 3rd year of their teaching careers in California. This represented 78 California BTSA programs in 107 school districts. However, the study was able to obtain survey responses from only 287 teachers, for a 26% response rate—most likely not representative. From the surveys, the study categorized teacher–respondents into high, middle, or low levels of induction engagement. The researchers then interviewed and observed smaller subsets of these teachers to obtain data for nine measures of teaching practice, such as instructional planning, reflection on practice, student questioning practices, feedback practices for students, and depth of student understanding. The study found that beginning teachers with high engagement in induction outscored the low-engagement group on seven of nine measures of teaching practice, although for only one measure were the differences at a statistically significant level. The authors concluded that, overall, their results demonstrated that BTSA/CFASST had a positive impact on teachers.

This is the only study we found that attempted to use multiple sources of data, including classroom observation, to measure teachers' practices, while sampling teachers from a wide variety of school districts and programs. However, along with a nonrepresentative sample, the study has weaknesses in the observation and interview data and processes, which the authors acknowledge and discuss. These included a lack of clarity regarding the definition of items, researcher fatigue problems handling the coding of observations on the same day they were collected, bias in the selection of students for interview, the unreliability of the insights of younger students, and the sheer number of items from the instrument.

The Effects of Teacher Induction on Student Achievement

We review four studies (see Table 1) that provide evidence about the relationship between beginning teachers' participation in induction and the academic achievement of their students. Two studies focused on California's BTSA program, one study examined a similar induction program in an unnamed large, urban, East Coast school district, and one study evaluated a similar induction program in New York City. Mentoring was the core element of these induction programs and hence the focus of these evaluations. Since all teachers in the samples participated in the mentoring program, these studies compared teachers according to the degree and type of support they received. The two studies in California and the study of an large, urban, East Coast district each found evidence that greater participation by beginning teachers in mentoring programs had a positive impact on their students'

achievement; the New York City study showed mixed effects—some positive effects, but also, in some comparisons, no effects.

One of these four studies is the project by Thompson and colleagues (2004), described above. In addition to examining the impact on beginning teachers' teaching practices, this study also examined the relationship between the degree of beginning teachers' engagement with district induction programs and their students' academic achievement. The researchers did not have access to data on gains over time in student achievement scores; instead, they used data on student achievement test scores at one point in time, limiting the study's ability to make conclusions about the impact of induction support on student achievement. Moreover, the study had a low response rate and a nonrepresentative sample because the analysis was able to obtain achievement test data for the students of only 144 of the 287 teachers who responded to the survey, reducing the sample to 13% of the target population of all third- to fifth-grade public school teachers in the 3rd year of their teaching careers in California. The study used hierarchical linear modeling techniques to examine the relationship between student test scores and each teacher's degree of induction engagement (high, medium, or low), after controlling for a number of key factors, including schoolwide academic performance, student socioeconomic status, and student English language-learner status, nested within individual teachers' classrooms. The analysis found that, across all six subtests of the standardized achievement exam, the students of teachers who had a high level of induction engagement outscored the students of teachers with a low level of engagement, after controlling for other factors. The authors concluded that although none of the score differences was statistically significant, the consistency of the results across all tests suggested that "BTSA/CFASST has a positive impact on student test scores" (Thompson et al., 2004, p. 13).

A pair of studies by Fletcher and colleagues also evaluated the effects on student learning of school district induction programs in California and in a large, urban, East Coast district. Fletcher, Strong, and Villar (2008) focused on the effects on student reading achievement of teachers' having different types of mentors. This study examined data from three California school districts. The district induction programs varied according to how they were implemented in the teachers' 2nd year. All three districts used mentors who were released from all teaching duties, with mentor to mentee caseloads of 1:15 in the 1st year. In the 2nd year, one district shifted to an in-school "buddy" mentor with no release time, one district doubled the mentor caseload, and the third district maintained the same caseload, thereby preserving the same high intensity of induction support. Using hierarchical linear modeling techniques, the researchers found that the third district, with a more intense mentoring model, showed higher class reading gains for its beginning teachers than the other two districts, after controlling for differences in district size, poverty, and student race/ethnicity. The authors could not infer causal relationships from this study because the limited sample size resulted in a design that did not let them distinguish school effects from district effects.

Another part of Fletcher et al.'s (2008) study focused on the third district, with its high-intensity mentoring model. Within each school, the analysis compared beginning teachers with veteran teachers as a whole. Veteran teachers may have had some induction support in the past, but they had not participated in the

district's comprehensive mentoring program. The objective of the analysis was to examine the impact of participation in mentoring on student test gains over 5 years. The analysis showed that although beginning teachers were more likely to be assigned to teach low-achieving classes, their students had, on average, equal or greater achievement than those of the more experienced teachers. A limitation of this design, comparing beginning to experienced teachers to test for effects of induction, is that the researchers did not know how much induction support the experienced teachers had received or to what extent more effective teachers might have moved to other, more attractive teaching positions or into school administration, thereby biasing the sample.

The second study by Fletcher and Strong (2009) compared two groups of beginning fourth and fifth grade teachers in a large, urban, East Coast school district. One group had support from a full-release mentor, whereas teachers in the other group were assigned a site-based mentor. The mentors received the same training, but they differed in caseload and release time. Teachers who received the support of a full-time mentor tended to have more low-achieving and low-income students than did teachers in the other group. In spite of this, students of teachers in the full-release mentor group showed greater achievement gains after 1 year. However, the opportunity to draw causal conclusions was again limited by the small sample size and a design that conflates potential teacher and school effects.

A final study (Rockoff, 2008) examined the effects of mentoring on student achievement (and on teacher retention) in New York City. As in the California studies, the investigator was not able to compare participating to nonparticipating new teachers, since all new teachers were enrolled in the district's program. The study compared beginning teachers to other newly hired teachers who had prior teaching experience and hence were not eligible for mentoring. Some of the latter may have had mentoring in prior schools, hence the comparison has limitations. However, within the group receiving mentoring, Rockoff (2008) compared those who received more time with a mentor to those who received less time.

Overall, the study found no differences in student achievement gains between newly hired inexperienced teachers who received mentoring and newly hired experienced teachers who did not receive mentoring. This is not unexpected. However, the study did find that teachers who received more hours of mentoring had higher student achievement score gains, in both math and reading, than those who had fewer hours of mentoring.

Since the activities of an induction program are at least one step removed from the students (see Figure 1), it is challenging to design research that can test the existence of a causal relationship between new teacher induction and student achievement. The above four studies show some consistency in results, but they also share a number of limitations, most of which the authors acknowledge. The most prominent weakness is that none of these studies involves random assignment of teachers to induction or mentoring groups. Neither students nor teachers are randomly distributed among classes and schools; parents may select school districts, schools, and even teachers; teachers are not randomly assigned among different levels of classes within schools; district resources may be differentially distributed among schools; classroom climates and other contextual conditions vary. All these factors may influence student performance and, unless controlled, may account for any differences in student achievement gains that appear to be the

result of teacher induction. With the possible exception of one small study using random assignment (Evertson & Smithey, 2000), this major limitation applies, in varying degrees, to all of the studies reviewed thus far for all three outcomes.

The Mathematica Study of the Effects of Induction on Beginning Teachers' Practices, Retention, and Student Achievement

The largest, most ambitious, and most important study investigating the impact of induction was funded by the U.S. Department of Education and conducted by a research team from Mathematica Policy Research of Princeton, New Jersey.[2] This study used a randomized controlled trial methodology. The major strength of a randomized controlled trial design is that it allows a study to isolate the impact of a treatment by ruling out other factors, such as the predispositions of participants and the character of the settings, that may affect the outcomes. This allows the researchers to make causal connections. We review this study separately, and at greater length, because of its size and importance and because it evaluated the impact of induction on all three sets of outcomes: beginning teachers' retention, classroom practices, and student achievement.

This study collected data from 1,009 beginning teachers in 418 schools in 17 large, urban, low-income public school districts. The sampled teachers were followed for 3 years, beginning in the 2005–2006 school year. Teachers' classroom practices were measured via classroom observations conducted in the spring of the 1st year, 2006. Data on teacher retention were collected via surveys administered in the fall of 2006, 2007, and 2008. Student achievement test scores were collected from district administrative records for the 2005–2006, 2006–2007, and 2007–2008 school years. This study randomly assigned the 418 schools to either the treatment or control condition, allowing for all new teachers in a school to be in the same group.

Beginning teachers in the treatment schools received "comprehensive" induction for either 1 or 2 years through programs offered by either Educational Testing Service (ETS) or the New Teacher Center, Santa Cruz (NTC). The programs included weekly meetings with a full-time mentor who received ongoing training and materials, monthly professional development sessions, opportunities to observe veteran teachers, and continuing evaluation of the teachers' practices. Beginning teachers in the control schools—those not assigned to receive comprehensive induction services—by default received the support normally offered to novice teachers by the district or school. The research design sought to ensure that the two teacher groups were balanced by race, gender, age, training, grade level, and certification.

The study's findings were mixed. For teachers' classroom practices, there were no significant differences between teachers in the treatment and control groups at midpoint in their 1st year on the job—the study did not assess impacts on practices past teachers' 1st year. For teacher retention, there were no significant differences between those in the treatment and control groups after each of the 3 years of follow-up. For student achievement, there were no significant differences between teachers in the treatment and control groups after either of their first 2 years. However, the study found that there were significant differences in the achievement of students of the teachers in the treatment and control groups in the teachers' 3rd year, based on the sample of teachers whose students had both pretest and posttest scores. These impacts were equivalent to moving the average student from

the 50th percentile to the 54th percentile in reading and to the 58th percentile in math. In other words, the study found that after 2 years of receiving induction, teachers' effectiveness significantly improved.

These results raise interesting questions. Some of the findings seem inconsistent with others from the study, some of the findings seem consistent with those in other studies, and some of the findings appear to contradict the findings of other studies. Given the size and importance of this study and its mixed findings, it is worth examining the study's characteristics, strengths, and limits in some detail. Later, in the conclusion, we return to the apparent consistencies and inconsistencies of findings within this study, and between this study and others, and try to summarize common ground and reconcile differences.

Differences Between Treatment and Control Groups

One issue concerns the degree, clarity, and consistency of differences between the treatment and control groups. The study documented that the intensity of induction support was greater in all aspects for the treatment group than for the control group, at a statistically significant level. This satisfies the starting assumption that the teachers in the treatment group were, in fact, receiving support that was more comprehensive than the baseline in the control group. But one of the key findings of the study was that induction and support are common, even in districts that supposedly did not have formal comprehensive programs. This is consistent with many of the earlier reviewed studies showing that induction is widespread. Moreover, this includes high-poverty schools, such as those sampled for this study. As reviewed earlier, an analysis of national data by Ingersoll and Smith (2004b) revealed that teachers in high-poverty schools are at least as likely as, if not more likely than, their counterparts in low-poverty schools to report they receive and participate in induction and mentoring.

As a result, as the authors carefully indicate, this study was not a comparison of those participating in induction to those not participating in induction. Nor was this study a comparison of those receiving formal induction to those participating only in some manner of informal induction. It was a comparison of teachers in schools that implemented a new "comprehensive" treatment based on two programs (from ETS or from the NTC) to those in schools that, for the most part, had formal induction programs already in place. Hence, this was not a study of the effects of getting induction per se, but a study of whether one type of induction—comprehensive—had different and better effects than the prevailing type of induction offered. This kind of comparison poses challenges and has implications for detecting effects.

The sampling design called for selecting districts in which the prevailing induction programs were not intensive, formal, or comprehensive. This would allow a distinct comparison when a subsample of schools in these districts then received the treatment of comprehensive induction. To obtain information on the degree of prevailing induction, the study interviewed district administrators and superintendents. One possible weakness with this approach is that it assumes that all schools in a district provide similar levels of induction to teachers and, moreover, assumes district-level officials are aware of the programs in particular district schools. However, individual school principals within a district could utilize school discretionary funds for the provision of a variety of supports, such as in-school mentors, orientation, professional development, release time, and professional learning

communities, resulting in within-district, cross-school variations in induction. And district officials may not be aware of these school-based efforts.

The study's descriptive data obtained from teachers reveal a different picture than that obtained from district officials. The data from teachers showed that, for some induction components, the control group support was not, in fact, greatly different from that provided to the treatment group. For example, 83% of control teachers reported having a mentor, compared to 94% of the treatment group. The ideal, of course, would be 100% participation by the treatment group, and far less by the control group. Likewise, average time spent with a mentor during the most recent teaching week was about 1.5 hours for the treatment group and about 1.25 hours for the control group. The average time spent one on one with a mentor was about 0.5 hours versus 0.2 hours, respectively. The average time observing and modeling lessons was 11 and 7 minutes, respectively.

Our point here is that if some of the control schools had induction services for beginning teachers that met, or came close to, the study's definition of comprehensive induction, the comparison between treatment and control groups would be muddied and the possibility of Type II errors would increase—acceptance of a null hypothesis of no differences in outcomes that, in fact, would be false. As a result, this kind of study could become the equivalent of a medical study that compares the effects of a specific dosage of a particular brand of aspirin to the effects of a variety of dosages of whatever other antipain medication the control group patients might have around the house, some of which could be similar to aspirin.

Variability within the treatment group also posed challenges. The comprehensive induction provided in the treatment group sought to closely follow the standard programs offered by ETS and the NTC but in some ways may have differed. For example, the mentors in the treatment group, although mostly having had prior mentoring experience, were all new to the two programs, whereas mentors in the study's control group were most likely working within a familiar program. Mentors' familiarity and experience with a program could be an important factor in success.

Variable participation in the treatment programs also occurred because not all teachers attended the five or six professional development sessions that were offered. Of those teachers enrolled in the ETS program, only 20% attended at least four of the five professional development sessions. Almost one third were present at two or fewer sessions. Likewise, of those enrolled in the NTC program, only 23% attended at least five of the six monthly sessions, and 22% missed at least three of the sessions. Participation in sessions was not mandatory, and it is unclear if nonparticipation was the result of a lack of motivation, a lack of confidence in the treatment, or problems with the implementation or provision of the treatment. Of course, ultimately, nonparticipation in a treatment has the same result as participation in an ineffective treatment. In both cases the treatment is not found to be successful. However, it is also worth understanding the reasons why a treatment was not successful. Nonparticipation in an otherwise effective treatment has different implications than participation in an otherwise ineffective treatment. To again use the above aspirin analogy, this could become the equivalent of a medical study that seeks to assess the effect of a specific dosage of aspirin, finds no effect, but also discovers that some of the participants took less than the specified dose of aspirin. It is unclear if the lack of effect is the result of not taking the full amount of aspirin or the aspirin's ineffectiveness.

This lack of clarity surrounding the degree and consistency of differences between the treatment and control groups has implications for the findings. On one hand, one might not expect to find large detectable differences in the outcomes for the treatment and control groups. On the other hand, it is striking that despite these issues, the study did find after 2 years significant differences in student achievement for those teachers getting comprehensive induction, compared to those getting the prevailing induction. In any event, it could have been the case that induction for both the treatment and control groups had a positive effect compared to getting no induction at all, but the study could not determine this because all got some induction.

The Measure of Teachers' Classroom Practices

A second issue concerns the outcome measure of teachers' actual classroom practices. Conducting and evaluating classroom observations of teachers in the field can be time-consuming, laborious, and expensive. As a result, such research often focuses on small samples. One important strength of this study is its relatively large teacher sample (1,009). But perhaps as a result of the large sample, this study used a relatively limited number of relatively short classroom observations of teachers done only in their first year of the study. Teachers were observed once during one reading and language arts lesson, in late spring during their first year of teaching, that is, after 6 or 7 months of treatment.

Regardless of how valid and reliable the observation instrument (the Vermont Classroom Observation Tool), it is unclear whether a single, relatively short classroom observation is sufficient to accurately characterize an individual's teaching strategies and classroom management or whether it is likely to detect differences between treatment and control teachers after about half an academic year. It is unfortunate that the study was not able to conduct multiple observations, especially including follow-up observations in the teachers' 2nd and 3rd years. This limits the ability of the study to discern later impacts and, in turn, what can be concluded regarding whether induction affects teachers' practices. It could be the case that the effects of comprehensive induction did not differ from those of the prevailing induction, or it could be the case that, like the delayed impact of induction the study found on student achievement, it would take more than one half year of participating in comprehensive induction before teachers' instruction techniques markedly improved over those receiving the existing induction.

Generalizability

A third issue concerns external validity and the issue of generalizability. The study focused on large, urban public school districts that had 50% or more students enrolled in the federal free or reduced-price lunch program for students from low-income families. From this group, the study included only districts for which district administrators reported low levels of existing induction and that were willing and able to participate, resulting in a sample of 17 districts. Large, urban, low-income school districts are the target of much attention and reform, and it is important to learn if induction can have a positive impact in such schools. But it is also important to recognize that the study sample was not representative of districts, schools, or teachers in the United States or of the subpopulation of large, urban, low-income school districts in the United States. This limits the ability to generalize from the study—it is unclear whether the results of comprehensive induction

found in the study's small sample of public school districts would hold true in other settings—a point to which we return in the conclusion.

In sum, the major advantage of a randomized controlled trial design is that it addresses threats to internal validity and allows the study to isolate the impact of a treatment and discern causal connections. However, it is unclear whether the advantages of the randomized design to detect impacts in this study have been partly undermined by other factors. Lack of full participation in the treatment by a portion of the treatment group, considerable levels of treatment experienced by teachers in the control group, limits in the outcome measure of teachers' classroom practices, and a nonrepresentative sample all pose possible limits to identifying differences in the effects of comprehensive induction compared to the prevailing induction and what we can conclude from this study's findings on effects.

Conclusions and Implications for Research

For decades researchers and commentators have called attention to the difficulties encountered by newcomers to elementary and secondary teaching, the lack of support provided to struggling novices, and their high levels of attrition during the first few years on the job (e.g., Johnson, 1990; Johnson & Birkeland, 2003; Lortie, 1975; Sizer, 1992; Tyack, 1974). Not all teacher attrition is, of course, negative; an early departure of a low-caliber teacher can be beneficial for the teacher, the students, and the school. But there is a growing consensus that high levels of teacher attrition, especially among beginners, are not cost-free. Teachers are an important resource, their production, training, and recruitment all entail costs, and the performance of newcomers improves if given sufficient time is not as high as that of veterans. As a result, in recent decades a growing number of states, school districts, and schools have developed and implemented induction programs for beginning teachers. The objective of these support programs is to improve the performance and retention of beginning teachers, that is, to enhance, and prevent the loss of, investments in teacher's human capital. In turn, there has been a growing body of empirical research designed to evaluate the effectiveness of these induction programs. The objective of this review is to critically evaluate this body of research.

As we have tried to point out in some detail, all of the studies reviewed have limitations and weaknesses of one sort or another. Despite these individual limits, however, the evidence collectively points in a similar direction. Overall, the studies we have reviewed provide empirical support for the claim that induction for beginning teachers and teacher mentoring programs in particular have a positive impact. Almost all of the studies we reviewed showed that beginning teachers who participated in some kind of induction had higher satisfaction, commitment, or retention. Likewise, for teachers' classroom practices, most of the studies reviewed showed that beginning teachers who participated in some kind of induction performed better at various aspects of teaching, such as keeping students on task, developing workable lesson plans, using effective student questioning practices, adjusting classroom activities to meet students' interests, maintaining a positive classroom atmosphere, and demonstrating successful classroom management. Finally, for student achievement, almost all of the studies reviewed showed that students of beginning teachers who participated in some kind of induction had higher scores, or gains, on academic achievement tests.

The major exception to this overall trend was the ambitious, large, and important randomized controlled trial conducted by Glazerman and colleagues (2010).

The results of this study were more mixed than most. This study did find that after beginning teachers had experienced 2 years of induction there were significant differences between the treatment and control groups in the achievement of their students. However, it also found no differences between the teachers in the treatment and control groups in their classroom practices in the 1st year and in teachers' retention over several years.

The study could not tell us whether the treatment and the control induction both had positive effects, or both had no effects on practices and retention, but simply that there were no significant differences in their effects on two of three outcomes. These mixed findings themselves are puzzling and seemingly contradict one another. Furthermore, finding a lack of effects on retention and classroom practices appears to sharply contradict most of the other studies we reviewed on those outcomes. This is significant because, in general, the research community views the results from randomized controlled trials as more reliable and valid than findings derived from other research designs (Riehl, 2006).

To further both research and policy it is, however, also important for us to not simply ignore conflicts among findings but to try to provide explanations to reconcile contradictory findings and also suggest future research needed to test such hypotheses.

One possible explanation for the conflicting findings regarding the effects of induction on beginning teachers' instructional practices could lie in differences in the duration of induction. The Glazerman et al. (2010) study found that it took time—at least 2 years of induction—for any differences in effects to show up in students' test scores. However, to examine the impact on their classroom practices, the beginning teachers in the sample were observed only once in the spring semester during their 1st year of teaching.

Notably, the five other studies on the effects of induction on classroom teaching practices all undertook multiple and lengthier classroom observations of each teacher in the study. Moreover, the largest of these five other studies observed the treatment group after they received induction for 2 years. Four of these five studies detected positive effects on teachers' practices; the fifth study had ambiguous findings. Hence, one explanation for the lack of effect on practices is that, like gains in student test scores, it could be the case that it takes more than a half year of participating in comprehensive induction before teachers' daily instructional practices visibly and consistently differ from those of teachers receiving the prevailing induction. This is consistent with the theory and rationale behind one of the comprehensive induction programs utilized in the Glazerman et al. study—the program offered by NTC. This model holds that on-the-job development of beginners takes more than 1 year, and hence beginning teachers in its program are required to receive 2 years of support (Moir, Barlin, Gless, & Miles, 2009).

Another possible explanation for the inconsistent findings regarding the effects of induction, especially on retention, lies in external validity—the issue of generalizability. Limits to the generalizability of findings from randomized controlled trials have been a point of debate in other fields. For instance, in medical research there has long been discussion among practicing physicians concerning the limits of results from clinical trials because patients in the field may differ from those enrolled in particular trials and trials may focus on population-level effects that are, by definition, overall averages (Chalmers, 1981; Riehl, 2006). The study by

Glazerman and colleagues intentionally sampled large, urban public school districts that had a majority of students from families below the federal poverty line. Although some of the other studies we reviewed similarly and solely focused on teachers in large, urban, low-income public school districts (e.g., Kapadia et al., 2007; Rockoff, 2008), most of the studies we reviewed did not. It is unclear whether the absence of effects of comprehensive induction on teachers' practices and retention found in the Glazerman et al. study's sample of large, urban public school districts would hold true in other types of districts.

That the effects of induction on retention can vary by setting is borne out by Ingersoll and Smith's (2004b) disaggregated analysis of national data. Their initial analysis of a national sample found that induction had strong positive effects on teacher retention (Ingersoll & Smith, 2004a). However, their follow-up analyses found that the impact of induction differed by school poverty level, with very strong effects in low-poverty schools and no effects in high-poverty schools (Ingersoll & Smith, 2004b). This latter finding is consistent with the findings in the study by Glazerman and colleagues. The Ingersoll and Smith data suggest that context matters and that induction's efficacy may depend on the school setting. Their hypothesis is that induction is not a panacea and that it, alone, may not be sufficient to reduce the high levels of teacher turnover that normally exist in many urban, low-income public schools. In other words, one explanation for the inconsistent findings regarding teacher retention is that although induction could, after a couple of years, positively affect teachers' practices and student achievement in high-poverty, urban public schools, nevertheless, receiving comprehensive induction as opposed to the prevailing induction alone may not be able to persuade teachers to stay in such schools at significantly higher rates.

This discussion on reconciling inconsistent findings and our review in general together suggest gaps in the research base and relevant questions that have not been addressed and warrant further research. We conclude by summarizing some of these below.

The Content of Induction

Much of the existing empirical research on the effects of induction is atheoretical; it examines what works, but not why or why not. A better marriage between the theory behind teacher development and the empirical research could advance our understanding. Future research could begin to clarify and sort out which elements, supports, and kinds of assistance are best and why. For instance, what should be the balance between induction focused on acquiring pedagogical skill versus that focused on subject-matter content?

Moreover, most of the existing research is uncritical as to the outcomes examined. Although the research has focused on an important set of outcomes (teacher commitment and retention, teacher classroom practices and student achievement), these do not exhaust the possible outcomes of induction. There are multiple and competing definitions of the goals of schooling and hence also multiple and competing definitions of the "effective" teacher. Definitions of the latter range from those teachers most able to engage students in higher order and critical inquiry, to those most effective at raising mature citizens, to those most sensitive to student diversity, to those most caring of children, to those best at promoting students' social and behavioral development, to those effective at raising student test scores.

It is convenient to assume that the "good" teacher is effective at most of the above tasks. But this may not be true. Indeed, coping with multiple and competing tasks has long been recognized as a central challenge for schools and teachers (Bidwell, 1965). Recent research suggests that teachers who are good at promoting some of the goals of education are not necessarily good at promoting other goals (see, e.g., Jennings, in press). Hence, it is important to ask which definition of the effective teacher is the goal for a particular induction program and if there are tough trade-offs. For instance, can an induction program simultaneously promote teachers' skill in engaging students in higher order inquiry and teachers' ability to teach standardized test taking, or are these contradictory imperatives calling for completely different induction emphases?

The Duration and Intensity of Induction

Both theory and some of the evidence suggest that the quantity of induction is important. That is, programs that are more comprehensive, or longer, or include more depth of support appear to be better. It is unclear, however, how long or intense induction programs need to be. Is there a minimal "tipping point" or threshold below which induction is of little value? On the other hand, is there an optimum program length and intensity for induction and mentoring programs, beyond which additional time invested diminishes in value? More specifically, is there an optimal quantity for particular components and activities. For instance, is there a significant difference in effectiveness depending on the amount of contact between new teachers and their mentors? Again, there is a role for theory in guiding the empirical research.

The Relative Costs and Benefits of Induction

Along with content and duration, induction programs also vary in their financial costs, and along with the question of which kinds and amounts of assistance are most effective is the question of which kinds and amounts of assistance are most cost-effective. Especially in periods of budget shortfalls, the "bang for buck" of such programs is, of course, crucial information to policymakers faced with decisions about which of many competing programs to fund. This is an area for which the research community could provide useful guidance to the policy community, but this is also an area for which there has been almost no empirical work done (for an exception, see Villar & Strong, 2007).

The Impact of Context

Existing research suggests that the content, duration, and costs of induction programs vary greatly among states, school districts, and schools. It is unclear, however, the extent to which the effects of and the cost-effectiveness of induction vary by setting. Are the content and duration of effective induction similar across settings? Or does induction need to be tailored to settings to be effective? Does effective induction in urban, low-income public schools necessarily differ from effective induction in suburban, affluent schools? Are some types and components of induction better for some types of teachers and students than for others? Does effective induction at the high school level differ from that at the elementary level? Moreover, are induction and mentoring programs particularly helpful for new teachers whose formal preparation is relatively weak, or are they helpful regardless of the quality of preclassroom preparation? Future research could illuminate these issues.

Notes

The order of the authors is alphabetical—each contributed equally.

[1] This 3-year project released an initial design report (Glazerman, Senesky, Seftor, & Johnson, 2006), annual reports of results after Years 1 and 2 (Glazerman et al., 2008; Isenberg et al., 2009), and a final overall report (Glazerman et al., 2010).

References

References marked with an asterisk indicate studies included in the critical review.

Achinstein, B., Ogawa, R. T., Sexton, D., & Freitas, C. (2010). Retaining teachers of color: A pressing problem and a potential strategy for "hard-to-staff" schools. *Review of Educational Research*, *80*, 71–107.

Bidwell, C. (1965). The school as a formal organization. In J. March (Ed.), *Handbook of organizations* (pp. 973–1002). Chicago, IL: Rand McNally.

Bradley, L., & Gordon, S. P. (1994). Comparing the ideal to the real in state-mandated teacher induction programs. *Journal of Staff Development*, *15*(3), 44–48.

Britton, E., Paine, L., Raizen, S., & Pimm, D. (2003). *Comprehensive teacher induction: Systems for early career learning*. Amsterdam, Netherlands: Kluwer and WestEd.

Chalmers, T. (1981). The clinical trial. *Milbank Memorial Fund Quarterly*, *59*, 324–339.

Charles, A. Dana Center. (2001). *Texas Beginning Educator Support System: Evaluation report for year one 1999–2000*. Retrieved from www.utdanacenter.org/downloads/products/txbess99_00.pdf

*Cohen, B., & Fuller, E. (2006, April). *Effects of mentoring and induction on beginning teacher retention*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

*Davis, B., & Higdon, K. (2008). The effects of mentoring/induction support on beginning teachers. *Journal of Research in Childhood Education*, *22*, 261–274.

*Duke, L., Karson, A., & Wheeler, J. (2006). Do mentoring and induction programs have greater benefits for teachers who lack preservice training? *Journal of Public and International Affairs*, *17*(2), 61–82.

Education Week. (2008). *Quality counts*. Arlington, VA: Editorial Projects in Education

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

Feiman-Nemser, S. (2001). From preparation to practice: Designing a continuum to strengthen and sustain teaching. *Teachers College Record*, *103*, 1013–1055.

Feiman-Nemser, S., & Schwille, S. (1999). *A conceptual review of literature on new teacher induction*. Washington, DC: National Partnership for Excellence and Accountability in Teaching.

Fideler, E., & Haselkorn, D. (1999). *Learning the ropes: Urban teacher induction programs and practices in the United States*. Belmont, MA: Recruiting New Teachers.

*Fletcher, S. H., & Strong, M. (2009). Full-release and site-based mentoring of elementary grade new teachers: An analysis of changes in student achievement. *New Educator*, *5*, 329–341.

*Fletcher, S. H., Strong, M., & Villar, A. (2008). An investigation of the effects of variations in mentor-based induction on the performance of students in California. *Teachers College Record*, *110*, 2271–2289.

*Fuller, E. (2003). *Beginning teacher retention rates for TxBESS and non-TxBESS teachers*. Unpublished manuscript, State Board for Educator Certification, Austin, TX.

Ganser, T. (1994). The perceptions of school principals about a mentoring program for newly hired urban school teachers. *Educator*, *30*(2), 13–23.

Ganser, T. (1996). What do mentors say about mentoring? *Journal of Staff Development*, *17*(3), 36–39.

Ganser, T. (2002). The new teacher mentors: Four trends that are changing the look of mentoring programs for new teachers. *American School Board Journal*, *189*(12), 25–27.

*Glazerman, S., Dolfin, S., Bleeker, M., Johnson, A., Isenberg, E., Lugo-Gil, J., & . . . Ali, M. (2008). *Impacts of comprehensive teacher induction: Results from the first year of a randomized controlled study* (NCEE 2009-4034). Washington, DC: U.S. Department of Education.

*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). Washington, DC: U.S. Department of Education.

*Glazerman, S., Senesky, S., Seftor, N., & Johnson, A. (2006). *Design of an impact evaluation of teacher induction programs* (Final Report No. 6137-070). Washington, DC: Mathematica Policy Research. Retrieved from http://www.mathematica-mpr.com/publications/PDFs/designimpact.pdf

Gold, Y. (1999). Beginning teacher support. In J. Sikula, T. Buttery, & E. Guyton (Eds.), *Handbook of research in teacher education* (2nd ed., pp. 548–594). New York, NY: Macmillan.

Gregson, J. A., & Piper, J. W. (1993). The Ohio induction process: Perceptions of beginning secondary school trade and industrial teachers. *Journal of Industrial Teacher Education*, *30*(2), 30–43.

Grissmer, D., & Kirby, S. (1987). *Teacher attrition: The uphill climb to staff the nation's schools*. Santa Monica, CA: RAND.

Grissmer, D., & Kirby, S. (1992). *Patterns of attrition among Indiana teachers, 1965–1987*. Santa Monica, CA: RAND.

Grissmer, D., & Kirby, S. (1997). Teacher turnover and teacher quality. *Teachers College Record*, *99*, 45–56.

Hafner, A., & Owings, J. (1991). *Careers in teaching: Following members of the high school class of 1972 in and out of teaching*. Washington, DC: National Center for Educational Statistics.

*Hahs-Vaughn, D., & Scherff, L. (2008). Beginning English teacher attrition, mobility, and retention. *Journal of Experimental Education*, *77*(1), 21–53.

Hegstad, C. D. (1999). Formal mentoring as a strategy for human resource development: A review of research. *Human Resource Development Quarterly*, *10*, 383–390.

*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*. Washington, DC: National Center for Educational Statistics.

Hobson, A. J., Ashby, P., Malderez, A., & Tomlinson, P. D. (2009). Mentoring beginning teachers: What we know and what we don't. *Teaching and Teacher Education*, *25*, 207–216.

Ingall, C. K. (2006). *Down the up staircase: Tales of teaching in Jewish day schools*. New York, NY: Jewish Theological Seminary Press.

Ingersoll, R.(2003). *Is there really a teacher shortage?* Philadelphia, PA: Consortium for Policy Research in Education, University of Pennsylvania. Retrieved from http://www.gse.upenn.edu/pdf/rmi/Shortage-RMI-09-2003.pdf

Ingersoll, R., & Kralik, J. (2004). *The impact of mentoring on teacher retention: What the research says*. Denver, CO: Education Commission of the States.

Ingersoll, R., & May. (2011). *Recruitment, retention and the minority teacher shortage*.Philadelphia, PA: Consortium for Policy Research in Education, University of Pennsylvania.

Ingersoll, R., & Merrill, L. (2010). Who's teaching our children? *Educational Leadership*, *67*(8), 14–20.

Ingersoll, R., & Perda, D. (2010). Is the supply of mathematics and science teachers sufficient? *American Educational Research Journal*, *47*, 563–595.

Ingersoll, R., & Perda, D. (2011). *How high is teacher turnover and is it a problem?* Philadelphia: University of Pennsylvania, Consortium for Policy Research in Education.

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

*Ingersoll, R., & Smith, T. (2004b, April). *The impact of induction and mentoring on beginning teacher turnover in high and low poverty schools*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

*Isenberg, E., Glazerman, S., Bleeker, M., Johnson, A., Lugo-Gil, J., Grider, M., & Dolfin, S. (2009). *Impacts of comprehensive teacher induction: Results from the second year of a randomized controlled study* (NCEE 2009–4072). Washington, DC: U.S. Department of Education.

Jennings, J. (in press). Beyond high-stakes tests: Teacher effects on other educational outcomes. In S. Kelly (Ed.), *Understanding teacher effects*. New York, NY: Teachers' College Press.

Johnson, S. (1990). *Teachers at work: Achieving success in our schools*. New York, NY: Basic Books.

Johnson, S., & Birkeland, S. (2003). Pursuing a sense of success: New teachers explain their career decisions. *American Educational Research Journal*, *40*, 581–617.

*Kapadia, K., Coca, C., & Easton, J. Q. (2007). *Keeping new teachers: A first look at the influences of induction in the Chicago Public Schools*. Chicago, IL: Consortium on Chicago School Research. Retrieved from http://ccsr.uchicago.edu/publications/keeping_new_teachers012407.pdf

Kyle, D. W., Moore, G. H., & Sanders, J. L. (1999). The role of the mentor teacher: Insights, challenges, and implications. *Peabody Journal of Education*, *74*(3–4), 109–122.

Leithwood, K., & Jantzi, D. (2009). A review of empirical evidence about school size effects: A policy perspective. *Review of Educational Research*, *79*, 464–490.

Lortie, D. (1975). *Schoolteacher: A sociological study*. Chicago, IL: University of Chicago Press.

Mitchell, D., & Scott, L. (1998). *The California Beginning Teacher Support and Assessment Program: 1998 statewide evaluation study*. Riverside: University of California, Riverside, Graduate School of Education, California Educational Research Cooperative.

Moir, E., Barlin, D., Gless, J., & Miles, J. (2009). *New teacher mentoring*. Cambridge, MA: Harvard Education Press.

Murnane, R., Singer, J., Willett, J., Kemple, J., & Olsen, R. (1991). *Who will teach? Policies that matter*. Cambridge, MA: Harvard University Press.

Perez, K., Swain, C., & Hartsough, C. S. (1997). An analysis of practices used to support new teachers. *Teacher Education Quarterly*, *24*(2), 41–52.

Riehl, C. (2006). Feeling better: A comparison of medical research and education research. *Educational Researcher*, *35*(5), 24–29.

*Rockoff, J. E. (2008). *Does mentoring reduce turnover and improve skills of new employees? Evidence from teachers in New York City* (Working Paper 13868). Cambridge, MA: National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w13868

*Roehrig, A. D., Bohn, C. M., Turner, J. E., & Pressley, M. (2008). Mentoring beginning primary teachers for exemplary teaching practices. *Teaching and Teacher Education*, *24*, 684–702.

Schaffer, E., Stringfield, S., & Wolfe, D. (1992). An innovative beginning teacher induction program: A two-year analysis of classroom interactions. *Journal of Teacher Education*, *43*, 181–192.

Scherer, M. (Ed.). (1999). *A better beginning: Supporting and mentoring new teachers*. Alexandria, VA: Association for Supervision and Curriculum Development.

Serpell, Z., & Bozeman, L. (1999). *Beginning teacher induction: A report on beginning teacher effectiveness and retention*. Washington, DC: National Partnership for Excellence and Accountability in Teaching. Retrieved from http://www.ericsp.org/digests/BeginningTeachInduction.htm

Sizer, T. (1992). *Horace's compromise: The dilemma of the American high school*. Boston, MA: Houghton Mifflin.

*Smith, T., & Ingersoll, R. (2004). What are the effects of induction and mentoring on beginning teacher turnover? *American Educational Research Journal*, *41*, 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*, 112–122.

Strong, M. (2009). *Effective teacher induction and mentoring: Assessing the evidence*. New York, NY: Teachers College Press.

Stroot, S. A., Fowlkes, J., Langholz, J., Paxton, S., Stedman, P., Steffes, L., & Valtman, A. (1999). Impact of a collaborative peer assistance and review model on entry-year teachers in a large urban school setting. *Journal of Teacher Education*, *50*, 27–41.

*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* (ETS-RR-04–31). Washington, DC: Educational Testing Service.

Tyack, D. (1974). *The one best system*. Cambridge, MA: Harvard University Press.

Villar, A., & Strong, M. (2007). Is mentoring worth the money? A benefit-cost analysis and five-year rate of return of a comprehensive mentoring program for beginning teachers. *ERS Spectrum*, *25*(3), 1–17.

Villeme, M. G., Hall, B. W., Burley, W. W., & Brockmeier, L. (1992). Are teachers receiving adequate support from their beginning teacher programs? The Florida experience. *Teacher Educator*, *28*(2), 10–16.

Waller, W. (1932). *The sociology of teaching*. New York, NY: Wiley.

Wang, J., & Odell, S. J. (2002). Mentored learning to teach according to standard-based reform: A critical review. *Review of Educational Research*, *72*, 481–546.

Wang, J., Odell, S., & Clift, R. (Eds.). (2010). *Past, present, and future research on teacher induction: An anthology for researchers, policy makers, and practitioners*. Lanham, MD: Rowman & Littlefield and Association of Teacher Educators.

Wang, J., Odell, S. J., & Schwille, S. A. (2008). Effects of teacher induction on beginning teachers' teaching: A critical review of the literature. *Journal of Teacher Education*, *59*, 132–152.

Wilson, S., Darling-Hammond, L., & Berry, B. (2001). *A case of successful teaching policy: Connecticut's long-term efforts to improve teaching and learning*. Seattle, WA: Center for the Study of Teaching and Policy.

Wollman-Bonilla, J. E. (1997). Mentoring as a two-way street. *Journal of Staff Development*, *18*(3), 50–52.

Youngs, P. (2007). District induction policy and new teachers' experiences: An examination of local policy implementation in Connecticut. *Teachers College Record*, *109*, 797–837.

Zey, M. G. (1984). *The mentor connection*. Homewood, IL: Dow Jones-Irwin.

Authors

RICHARD M. INGERSOLL is professor of education and sociology at the University of Pennsylvania, Graduate School of Education, 3700 Walnut St., Philadelphia, PA 19104; e-mail: *rmi@gse.upenn.edu*. His research is concerned with the management and organization of elementary and secondary schools and the character and problems of the teaching occupation.

MICHAEL STRONG was director of research for the New Teacher Center at the University of California, Santa Cruz; e-mail: *mastrong@ucsc.edu*. His current research focuses on the assessment of teacher effectiveness.