

SRI International

February 11, 2004

Review of Research on the Impact of Beginning Teacher Induction on Teacher Quality and Retention

ED Contract ED-01-CO-0059/0004
SRI Project P14173

Prepared for:

Melanie Rogers
U.S. Department of Education
Institute of Education Sciences
National Center for Education Evaluation and Regional Assistance

Submitted by:

Alejandra Lopez
Andrea Lash
Monika Schaffner
Patrick Shields
Mary Wagner

Acknowledgments

We thank the following individuals for their assistance in producing this review: Roneeta Guha, Cimone Satele, Heidi Skolnik, Yukie Toyama, and Andrew Wayne.

Review of Research on the Impact of Beginning Teacher Induction on Teacher Quality and Retention

In reaction to concerns about teacher quality and the stability of the teacher workforce, policy-makers have increasingly initiated programs designed to support the entry of new teachers into the profession. Between the 1990-91 and 1999-2000 school years, the proportion of beginning public school teachers who participated in an induction program increased from 51 percent to 83 percent (Smith & Ingersoll, 2003). Such programs typically include the assignment of a more experienced mentor to a new teacher and may involve a series of other supports, such as orientation meetings, time to meet with colleagues, assessment, and targeted professional development. In practice, induction programs vary widely across—and even within—schools and districts, ranging from single events to regular and highly structured instructional and classroom management support (Fideler & Haselkorn, 1999).

Regardless of their scope or structure, induction programs share a simple logic: because new teachers tend to be less effective and are more likely to leave the profession than their more experienced peers, targeted support should be provided to orient them to the profession and to assist them to learn their craft. The goals are to make new teachers more effective earlier in their careers and to keep them in the profession.

Yet, as with many educational reforms, the enthusiasm of advocates and policy-makers for induction programs may be based more on intuition and anecdotes than on strong research evidence. Consequently, the National Center for Education Evaluation within the Institute of Education Sciences has supported SRI International to explore the feasibility of conducting a randomized trial to assess the impact of induction programs on teacher retention and teacher quality. As a first step in that process, SRI undertook a thorough review of the empirical evidence on the effectiveness of induction programs.

The main goals of the review were to learn what the existing literature tells us about (1) whether induction programs affect teacher retention and teacher quality (particularly in terms of student achievement) and (2) which components of induction programs are the most promising in terms of improving teacher retention and teacher quality (again, particularly in terms of student achievement). The review focuses on induction programs for *beginning* teachers, defined as being in their first or second year of teaching, though some of the studies we review used broader definitions that included teachers new to the state, district, or school. The bottom line of the review is that in spite of the intuitive appeal of induction programs and a plethora of literature on what induction programs should encompass, there are few examples of rigorous research that demonstrate the impact of induction efforts.

The document is structured as follows. We begin with an explanation of how we conducted the literature review. We then present an overview of the findings. In the third and fourth sections of the document, we review the studies individually, discussing first those related to teacher retention and then those that focused on teacher quality. We conclude with a summary of findings from the research review. Detailed descriptions of the studies are included in Appendix A.

The Literature Search

Our objective in this review was to summarize and critique empirical research on the impact of beginning teacher induction on teacher retention and teacher quality (particularly studies in which teacher effectiveness was evaluated by using student achievement measures). With a focus on experimental and quasi-experimental research written since 1980,¹ we conducted preliminary searches in a number of social science databases to assess the availability and

¹ We chose to limit the search to publications written since 1980 because work older than that would be less likely to represent classrooms, teaching, and induction programs today.

prevalence of such literature. We chose to focus on the Educational Resources Information Center (ERIC) and PsycInfo as primary sources because of the lack of relevant articles in the other databases. Our initial search yielded 209 publications, using the following ERIC descriptors: (beginning teacher induction or teacher induction) and (teacher retention or teacher quality or teacher effectiveness or student achievement). On the basis of an initial review of the 209 abstracts, we expanded the ERIC search to include additional descriptors: (beginning teacher induction or teacher induction) and (program effectiveness or program evaluation). This second ERIC search yielded 170 publications not picked up initially.

To determine which publications fit the criteria for inclusion in this literature review, all of the studies' abstracts were divided among four readers, who classified each publication into one of three groups (yes, maybe, no) based on the likelihood that it was an empirical study of the impact of teacher induction. All four readers then read the abstracts for the articles that were *not* clearly outside the scope of the study (i.e., readers examined the yes and maybe groups) and rated them on a scale of 1 to 6, detailed in Exhibit I. Where there was disagreement among raters, two readers read the full article to decide a rating. Full copies of the experimental, quasi-experimental, and correlational studies (i.e., publications rated as a 1, 2, or 3) were obtained and read to verify the rating. In addition, existing literature reviews on the topic (i.e., reports from the databases and other reviews recommended by colleagues, all categorized as 5 in Exhibit I) were gathered and read, and seven additional studies were identified that had not shown up in our database searches. As presented in Exhibit I, this raised the total number of articles identified for examination to 387.

In addition to the research databases and previously published literature reviews on the topic, we also browsed a number of organizations' Web sites in search of relevant reports;

Exhibit I: Summary of Documents Evaluated (N=387)

<i>Rating of Publication for Inclusion in Literature Review</i>	<i>Number of Documents</i>
1. Experimental designs that call for a random method to be used to assign participants to treatment and comparison groups.	3
2. Quasi-experimental designs that include a planned comparison between groups of people (e.g., treatment and control), between occasions of measurement (e.g., pretest and posttest), or both.	9
3. Quasi-experimental designs having only one group and one occasion of measurement, and correlational studies that examine associations among variables within a single population.	32
4. Qualitative research.	22
5. Reviews of research on beginning teacher induction.	23
6. Not empirical and/or not a fit with the topic of interest.	296
Unrated because abstracts not available for review.	2

however, the documents located were all policy briefs with purely descriptive information, primarily about the purpose of induction efforts or state policies and programs.² Other, more general Internet searches were also conducted, primarily using keywords similar to the ERIC descriptors mentioned earlier, but they did not identify additional research reports for review.

In the end, 12 experimental and quasi-experimental works were selected for review. Appendix B lists the references for all reports that are experimental, quasi-experimental, correlational, or reviews of literature.

Overview of the Research Reviewed

Here we provide an overview of the research we located on the impact of induction on beginning teachers in terms of retention and teacher quality. Of the 12 studies reviewed, six are evaluation

² Organization Web sites browsed include: Educational Testing Service; National Education Association; American Federation of Teachers; American Institutes for Research; RAND; Westat; WestEd and other regional labs; Abt Associates; and the Center for Research on Educational Equity, Assessment, and Teaching Excellence.

reports examining specific induction programs, three are conference papers, two are peer-reviewed articles, and one is a book chapter. The work was produced between 1985 and 1994, with the exception of one 2002 publication. Only three of the studies randomly assigned beginning teachers to control and treatment groups³; and even these provided limited, if any, information about how the randomization was carried out or about attrition during the research. The other nine studies are quasi-experimental, including planned comparisons between groups of beginning teachers or pre and post comparisons of one group of beginning teachers.

To help us evaluate the rigor of the research, we developed an analytic matrix to keep track of quality indicators and other key information about each study (see Appendix A). In creating this analytic matrix, we took into account guidelines specified in the *Study Design and Implementation Assessment Device*, a resource developed by the What Works Clearinghouse (available online at <http://www.w-w-c.org/standards.html>) to assess the strength of research designs in studies investigating causal relationships. Areas we evaluated included quality of the treatment(s), quality of the outcome(s), adequate reporting of statistical tests, accurate estimates of effect sizes, comparability of the treatment and control groups, possible confounding factors, and the generalizability of findings.

The majority of the research we reviewed are evaluations of specific induction programs (including California's Mentor Teacher Induction Program, Toronto's Peer Support Pilot Project, the Texas Beginning Educator Support System, the Florida Beginning Teacher Program, New York City's Mentor Teacher Internship Program, Alberta's Initiation to Teaching Project, and Indiana State University's Project CREDIT: Certification Renewal Experiences Designed to Improve Teaching). The other two studies are research that investigated different approaches to

³ In one of the studies, Gold (1987), principals were responsible for the random assignment of mentors to beginning teachers, so it is questionable whether this research implemented a truly experimental design.

an induction activity, mentoring: one examined a formal versus informal system of mentoring, and the other examined the training of mentor and mentee pairs in classroom management. Mentoring was a defining feature of all but one of the induction treatments researched in this body of work, in most cases used in conjunction with other activities, such as workshops and seminars, the development of professional development plans and portfolios, release time for professional development, classroom observations with feedback, or classroom management training. The quality of the induction “treatments” was difficult to assess since evaluators and researchers seldom reported on their implementation.

The studies included one or both of our outcomes of interest: retention and teacher quality. Retention was measured by counts of teachers remaining in their positions and teachers’ self-reports about plans to continue teaching. Teacher quality was measured in terms of student performance, teacher competencies and knowledge base, use of class time and classroom management, and attitudes/behaviors/morale. Typically, researchers used only one measure of each outcome variable, and, in many cases, the measures were weak (e.g., a single questionnaire item) and/or information about their quality was not reported. The studies related to retention tended to rely on teachers’ self-reports about their plans to remain in teaching, as opposed to actual counts of teacher retention. Had there been more studies that focused on actual counts of teacher retention, we would not have included in this review those that relied solely on teachers’ self-reported plans. Of the studies on teacher quality, one considered student achievement data; the others often relied on classroom observations to assess the impact of induction on beginning teachers. Had there been more research focused on student achievement outcomes, we could have focused our efforts on that body of literature.

In the set of studies on retention (n=6), three reported a positive relationship between participation in a teacher induction program/activity and teachers' staying in the same teaching position and/or the teaching profession (two of which lacked tests of statistical significance), two of the studies presented mixed results (one of which lacked significance tests), and one reported no impact. Of the reviewed research on teacher quality (n=10), four reported a positive relationship between participation in an induction program/activity and beginning teacher effectiveness (one of which lacked significance tests), four studies indicated mixed results, and two found no impact.

None of the research we examined included estimates of effect sizes. A number of the studies did not conduct tests of statistical significance on reported comparisons between treatment and control groups, and of those that did, some did not report the test statistics or only summarized them in the text of the document. In addition, some of the questionnaire-based studies had low response rates—e.g., in the New York City Board of Education study, only 41 percent of the treatment group and 22 percent of the control group responded; the rest of the studies with surveys did not present or discuss response rates, though we can tell they are low in the ones where N's are provided. The set of reviewed research is weak in these respects, making it difficult to assess, and draw conclusions about, the impact of teacher induction on beginning teacher quality and retention.

The most common sources of potential confounding in this collection of studies include a lack of comparability of the treatment and comparison groups, the possibility of treatment contamination due to the presence of treatment and comparison groups in the same school, problematic timing of the treatment, weak control of the measurements gathered, failing to track teachers in an induction program spread out across multiple regions, and poor attention paid to

differential attrition rates of the study groups. Limits to the generalizability of the findings include problematic definitions of beginning teacher participants, lack of information about the methods used to select beginning teachers for the study, and insufficient descriptions of the participants and of how well the study participants represented the population of interest.

Research on the Impact of Induction on Teacher Retention

Six of the 12 studies reviewed examined the impact of teacher induction on retention, all using a posttest-only design with comparisons of beginning teachers. This body of research is primarily quasi-experimental (i.e., with no random assignment of teachers to treatment and comparison groups) and based on self-reports of plans to continue teaching. Only two of the studies on retention analyzed records kept by the state or school district, Dana Center (2002) and Gold (1987). These will be discussed first, followed by the four studies that measured retention based on beginning teachers' plans to remain in the profession (New York City Board of Education, 1993; Cheng & Brown, 1992; Gunter, 1985; Brown & Wambach, 1987).

In their evaluation of the Texas Beginning Educator Support System (TxBESS), the **Charles A. Dana Center (2002)** examined retention rates of three cohorts of beginning teachers in Texas (and the relationship between student performance and teacher characteristics, including participation in the induction program—though this part of the study will be discussed in the next section on teacher quality research). Participants in the TxBESS program were provided instructional and mentor support as well as formative assessment during their first year of teaching in Texas public schools. School districts were supported by educational service centers, located in each of the state's 20 regions, responsible for tailoring the TxBESS program to meet the needs of the local districts. The program, which began in 1999-2000, served 12 percent of new teachers in the state in its first year of full implementation, 2001-02. In the first year of the

evaluation (1999-2000), 998 beginning teachers were included in the study; in 2000-01, 2,059 teachers; and in 2001-02, 3,058 teachers. Retention rates over a three-year period were compared for TxBESS participants (the treatment group) and nonparticipants (the comparison group). Overall, TxBESS teachers showed higher retention rates, especially for minority groups and teachers at the high school level (though no tests of statistical significance were reported for these findings): 87 percent of African-American teachers who participated in TxBESS returned for a second year of teaching, compared with 77 percent of those who did not participate in the program; 91 percent of Hispanic teachers who participated in TxBESS returned for a second year of teaching, compared with 73 percent of those who did not participate; 89 percent of TxBESS teachers at the high school level returned for a second year of teaching, compared with 79 percent of those who did not participate.

There are several limitations in this study, one being that some of the districts had no teacher representation in the comparison group since they required all of their new teachers to participate in the TxBESS program. The authors also expressed some concern about the inability to track all of the program participants in the Texas Education Agency Public Education Information Management System (TEA-PEIMS) database since names for the induction program varied across some of the regions in the state. To some extent, the implementation of the program also varied across the state (e.g., whether participants were assigned a mentor, the frequency of their interaction with a mentor, and the degree to which mentors were matched with mentees on subject taught). In terms of the comparability of the treatment and comparison groups, data presented in the research indicate similar demographic compositions of the two groups of beginning teachers, though it should be noted that participants in the TxBESS program tend to include a greater percentage of teachers who work in economically disadvantaged areas.

The **Gold (1987)** study also examined the impact of mentoring on beginning teacher retention, based on counts of teachers who stayed in the district. Gold evaluated a New York City mentor program in which retired teachers served as consultants to new teachers in schools with high levels of teacher turnover, with the goals of improving teaching and reducing the attrition of first-year teachers. The larger study focused on mentor training and the types of assistance that mentored and nonmentored teachers received in their first year on the job, as measured by responses to post-program questionnaires. We will review here only the portion of the evaluation that examined the program's impact on retention.

This is the only retention study included in our review that used an experimental design with random assignment of beginning teachers to treatment and control groups; each mentor was assigned to a school, and then the principals randomly assigned three new teachers to each mentor and three other new teachers in the same school to a control group. Retired teachers interacted with beginning teachers for 66 hours during the course of the school year. The mentors were selected from a pool of retirees, recommended by principals, who met certain criteria, including a minimum of six years of teaching experience, a positive attitude toward teaching, strong communication skills, and an ability to work well one-on-one. Additional selection criteria included previous experience in a helping role in the teaching profession (e.g., as a supervisor of student teachers) and completing an acceptable written critique of a videotaped classroom lesson. During the summer prior to the program, mentors received 18 hours of training, designed to clarify the mentor's role and to help transition them confidently into this position; in addition, three half-day sessions were held during the school year for mentors to share ideas and problem-solving strategies. The researcher reported that most of the mentors' time was spent on instructional issues: observing the new teachers teaching, working in the

classroom, modeling effective teaching techniques, and meeting with the new teachers to plan lessons, prepare materials, or provide feedback.

The retention part of the evaluation was based on a sample of 160 mentored teachers and 113 control group teachers (unmentored). One drawback in the reporting of this research is that the method used to measure attrition was not explained, nor was the definition of attrition specified (e.g., leaving the profession vs. leaving the district). Results indicate that attrition was lower among beginning teachers in the mentor program (an average of 1.88 percent attrition for mentored teachers, compared with 3.54 percent for control group teachers in the same school), though no tests of statistical significance were conducted. The author also reported that the attrition rate for teachers in comparable schools where no mentoring took place was 4.54 percent. The author reported that these attrition rates may have been low overall because a substantial increase in the salary schedule for new teachers was implemented during the year of the evaluation.

One limitation of the study is that no information was provided to assess whether the assignment of teachers to groups within schools was truly random since principals at individual school sites were responsible for the assignments. In addition, the author indicated that reassignments had to be made in some cases when a teacher was resistant to accepting assistance from the mentor assigned. Other weaknesses of the study include that mentored and control group teachers were in the same schools, and that the schools that participated tended to be the ones with principals who were the most eager to have mentors and the ones that had the highest rates of attrition for new teachers.

Unlike the Dana Center and Gold studies, which focused on state or district records, the other four retention studies measured retention in terms of self-reported plans to continue

teaching. The **New York City Board of Education (1993)** evaluated whether participation in the Mentor Teacher Internship Program improved retention (as well as teaching effectiveness and socialization into the school community, which will be discussed in the next section on teacher quality research) during the 1992-93 school year. In this induction program, mentors received training and release time and were matched with newly hired uncertified teachers; interactions included peer coaching, the opportunity to take a graduate-level course provided by the United Federation of Teachers, and other ongoing professional development activities over the school year. The graduate courses available to teachers focused on teaching techniques and classroom management, specific content areas, or content or technique combined with a focus on specific student populations. In a quasi-experimental design, 1,264 interns were surveyed as the treatment group, and 100 teachers without a mentor were surveyed as the comparison group; at the end of the school year (May), both groups were asked about their plans to remain in the teaching profession. Overall, the mentored interns were more somewhat more likely than the nonmentored new teachers to report that they would remain in the teaching profession for the next five years (64 percent versus 55 percent), though no test of statistical significance was reported. This was especially the case if the mentor was in the same subject and/or grade level as the new teacher: 70 percent said they planned to be teaching for five years, compared with 57 percent of the mentored teachers whose mentor did not match their subject or grade level (statistical significance was reported). Another finding (reported as statistically significant) is that retention was higher for program participants who were assigned mentors at the beginning of the school year rather than mid-year (84 percent versus 77 percent).

The authors described the treatment and comparison groups in terms of degree attainment, teaching experience, certification, and grade level taught; though they likened the

groups on a number of characteristics, it is not clear from the data presented that these two groups are comparable. For example, a larger percentage of treatment group teachers than of comparison group teachers had not taught before and were uncertified. Another limitation to the research is that not all of the program participants were assigned a mentor at the same time—some received one at the beginning of the school year, others in mid-year (January). Also, the interns had different levels of teaching experience—i.e., some of the teachers had taught the year before, even though they were uncertified at the time. In addition, the fact that uncertified teachers were the study participants might limit the generalizability of the findings to other types of teachers.

Cheng and Brown (1992) examined the success of a formal induction program, Toronto's Peer Support Pilot Project, during two years of implementation, 1990-91 and 1991-92. New teachers in the program participated in workshops and professional development activities, including an orientation meeting (both years); meetings in October and May to share common concerns (1990-91); a series of three half-day workshops in November, February, and April (1991-92) on themes identified by participants as important for the professional development of new teachers (e.g., whole language learning); and five days of release time for professional dialogue and sharing (both years), including the creation of individual professional development plans. Program participants also had mentors, though no information is provided about the frequency and nature of their interactions, only that they were matched by grade level or program within the same school. Included in the evaluation were 25 pairs of teachers (beginning teacher and mentor) as the treatment group during the first year and 38 pairs during the second year; comparison groups of beginning teachers were also included, 30 in 1990-91 and an unspecified number in 1991-92.

In terms of retention, self-reports were gathered at the end of the school year from the first-year teachers about the decision to become a teacher, choosing teaching as a career again, and staying in the teaching profession. During the first year of the evaluation, all of the pilot program teachers felt that they had made the right career choice; almost all said that they would make the same decision again, whereas fewer in the comparison group said the same; and a larger percentage of pilot project teachers (76 percent) expected to still be teaching five years later, compared with nonparticipant beginning teachers (60 percent). In the second year of the evaluation, almost all teachers in both groups thought that they had made the right career choice and would make the same decision again, and slightly more teachers in the pilot project (97 percent) than in the comparison group (91 percent) expected to remain in teaching two years later. One limitation of the research is that no tests of statistical significance were presented.

Unlike many of the other studies we are reviewing, Cheng and Brown provided descriptive data about the treatment and comparison groups included in their research, though the description lacks a clear picture of the population of *all* first-year teachers in Toronto schools, the target group for this induction program. In the first year of the evaluation, the demographics of the treatment and comparison groups appear to be comparable; however, a potentially confounding factor is that teachers who applied to participate in the induction program but did not meet the criteria for selection⁴ were assigned to the comparison group. In the second year of the study, a larger proportion of teachers in the comparison group than in the treatment group had prior teaching experience. Another potential limitation is that the comparison groups were defined differently in each year of the evaluation. In the second year, since all interested

⁴ In the first year of the evaluation, the selection of participants for the Toronto Peer Support Program was based on the following: that the new hire be an inexperienced teacher, that a volunteer mentor be available in the same school as the new hire, and that the experienced teacher and new hire be teaching a similar grade level or program. Also, the selection process had to satisfy the goal of including as many schools as possible in the project.

participants were accepted into the program, a randomly selected group of elementary teachers was assigned to be the comparison group (unlike the first year where the comparison group was made up of applicants who were not selected for the program).

Also using teacher self-reports to measure retention, **Gunter (1985)** examined the impact of the Florida Beginning Teacher Program on participants as compared with nonparticipants. The research question of interest in this study was whether first-year teachers who participated in the induction program differed in attitude, behavior, and/or competence from first-year teachers who did not participate in the program. Seventy-seven matched pairs were included in the study from three districts in the state (one teacher participated in the program; the other did not); they were matched in terms of age, race, sex, grade level taught, subject area taught, degree, and university attendance (in-state or out-of-state). Program participation was based on the date when teachers applied for certification: those who applied before a new law went into effect were not required to participate; those who applied after the law was enacted were subject to the new requirement to participate in Florida's Beginning Teacher Program. It is not clear how, if at all, these two groups of first-year teachers might have differed. Rather than being assigned a mentor (as has been the focus in the studies reviewed thus far), beginning teachers in the treatment group of this study were assigned a support team consisting of an administrator, a peer teacher, and another educator, who created a professional development plan for them; participant teachers also experienced classroom observations with feedback from each member of their support team, and assistance creating a portfolio. The amount and quality of other training that inductees received depended on how the program was implemented locally.

Gunter's research indicates mixed results regarding the impact of Florida's Beginning Teacher Program on the retention of new teachers. The retention part of the study is based on

first-year teachers' survey responses, as well as teacher absentee rates. The number of teachers who planned to remain in the *profession* indefinitely did not differ significantly for participants (67 teachers) and nonparticipants (62 teachers); however, the number who reported plans to remain in the same *position* indefinitely differed significantly between groups (36 participant teachers versus 46 nonparticipants). In terms of absentee rates, there was no statistically significant difference between first-year teachers who took part in the induction program and those who did not.

One possible contaminating factor in this research is that treatment and comparison group teachers could be in the same school. Also, it should be noted that the comparison group teachers, though they did not participate in any formal induction program, had regular district evaluations two to three times per year, making it difficult to isolate the impact of observations through the formal induction program from that of evaluations conducted as part of a larger district effort.

Brown and Wambach (1987) evaluated the impact of California's Mentor Teacher Induction Project on beginning teachers, including student teachers and first-year teachers. The program emphasized contact with a mentor, opportunities for both formal and informal professional support, and the involvement of mentor teachers in seminars. The study included treatment and comparison groups of both populations, student teachers and first-year teachers in California, though no additional information was provided about sample size or sample characteristics. Student teachers in the treatment group were assigned a mentor teacher for a seven-week student teaching experience, including their mentors' attending three of seven weekly seminars held for the student teachers. First-year teachers in the treatment group were matched for one year with a mentor who had a similar teaching assignment in terms of subject

and grade level taught; they received support including ongoing contact to discuss problems and concerns, classroom visits by the mentor, visits to the mentor's classes, involvement in program seminars (no specifics were given in the report about the content of these seminars), and monthly support sessions with professional counselors and university faculty. In both treatment groups, the new teachers worked with mentors who were experienced teachers and had applied or been recommended to serve as mentors for the programs, though no details were provided about their training; the mentors in both programs also received a small stipend for their time.

At the end of the school year, student teachers and first-year teachers completed questionnaires on which they were asked, "Will you continue teaching?" and provided with four response options: no, unsure, yes—probably, yes—definitely. The authors found no significant differences between treatment and comparison groups in each population in teachers' intent to stay in the profession. Significance tests were conducted in this research, though test statistics were not presented (a p-value was reported with regard to only one comparison). Since no information was provided about the sample of teachers who took part in the study, we cannot assess the comparability of treatment and comparison groups of first-year teachers and student teachers, nor can we evaluate the generalizability of the findings.

Research on the Impact of Induction on Teacher Quality

Ten of the 12 studies reviewed considered the impact of beginning teacher induction on teacher quality, measured in terms of student performance, teacher competencies and knowledge base, use of class time and classroom management, and attitudes/behaviors/morale. Five of the publications used a posttest-only design, four used a pretest/posttest design, and one used both. Two of the 10 studies on teacher quality were experimental; the rest were quasi-experimental. Unlike the research on retention, which relied heavily on data collected via self-reports from

beginning teachers, the research on teacher quality also considered classroom observations (Klug & Salzman, 1991; Stallion, 1988; Schaffer, Stringfield, & Wolfe, 1992; Ratsoy et al., 1987; Gunter, 1985; and Kilgore & Kozisek, 1989, to some extent) and student achievement data (Dana Center, 2002). Only three of these studies relied solely on teacher self-report data (New York City Board of Education, 1993; Henry, 1988a; Cheng & Brown, 1992). The observation-based studies will be reviewed first, followed by the ones based on teacher self-report and student achievement data. Four of these studies also examined the impact of induction on retention, so descriptions in this section will focus on aspects of the research related to teacher quality that have not yet been reviewed.

Klug and Salzman (1991) compared the impact of two induction approaches on teacher competencies, attitudes, and morale among a group of first- and second-year teachers, randomly assigned to one of two induction approaches. Participants in this experimental study included beginning teachers (in their first or second year of teaching) from three rural or small-city school districts in southwest Idaho. All of the teachers were assigned a mentor, though it is not clear in the paper whether they were matched on any criteria. Group 1 (strong induction program) teachers received classroom observations by the mentor team two hours per month during the first semester, semimonthly observations during the second semester, and two team meetings per semester; overall, these beginning teachers spent about 72 hours with mentor teachers over the course of the year. Group 2 (weak induction program) teachers underwent informal mentoring via a “buddy” system, with no structured number of hours of contact specified and no guidelines given about the nature of mentor-mentee interactions that should take place. In this research, pre- and posttest data were compared for both treatment groups to examine the impact of a strong

versus a weak induction program on teacher quality. Inadequate information was provided to assess the comparability of the two groups of teachers.

Mentors rated beginning teachers' competencies, based on videotaped classroom observations, using two of the five scales on the Teacher Performance Assessment Instrument (TPAI): classroom procedures (i.e., instructional practices within the classroom setting) and interpersonal skills (i.e., ability to create a comfortable social setting, to demonstrate warmth and friendliness, and to manage classroom interactions). In addition, beginning teachers completed a questionnaire, the Purdue Teacher Opinionnaire, on which they responded to 145 statements about morale and attitudes toward teaching. Unlike many of the other studies reviewed here, Klug and Salzman provide a thorough description of the validity and reliability of both the TPAI and the questionnaire they used in their research. Reliability indices of the TPAI ranged from 0.85 to 0.93, and the overall reliability coefficient for the Purdue Teacher Opinionnaire was 0.96.

Overall, this study offers mixed results (with reports of statistical significance) regarding the impact of induction on teacher quality. There was no significant difference between the two groups of teachers in the ratings of teacher competencies; both groups had slight declines. In terms of questionnaire responses, Group 1 (strong induction program) indicated greater positive increases than Group 2 (weak induction program) in five areas: rapport with the principal, curriculum issues, teacher status, community support, and community pressures and no differences in change of opinion about teacher salary and school facilities.

The authors expressed concern that teachers might have been uncomfortable with the videotaping of their teaching, and hence the competency ratings may not have been a strong gauge of their teaching abilities. Also problematic, the authors noted that qualitative data from the same study suggest that teacher competencies increased over the course of the year,

contradicting ratings based on the videotaped classroom observations. In addition, the extent to which the rural and small-city settings where the research took place may limit the generalizability of the findings was not addressed.

Using an experimental research design, **Stallion (1988)** assessed the effects of classroom management training and mentoring on beginning teachers (also referenced in Stallion & Zimpher, 1991). The training was embedded in a larger school-university teacher induction program, though other aspects of the program were not evaluated. Thirty-five pairs of beginning teachers and mentors (matched by their principals on grade level or subject taught) participated in the study and were randomly assigned to one of three groups, stratified by grade level and subject taught: Group 1—both the mentor and mentee participated in the classroom management training; Group 2—only the mentor participated in the training; Group 3—neither teacher participated. The number of teacher pairs assigned to each of the three groups was not clear. Mentees who took part in the research were teachers new to their classrooms and included teachers returning from leave, those at a new grade level or in a new subject area, and those new to their school or district. This is a broader definition of beginning teachers than used in much of the other research we reviewed, and the authors note that it may have been problematic in that some of the teachers who had teaching experience in another school, district, grade level, or subject area, may have resented having to take part in the program geared to novice teachers.

The classroom management training was research-based, and, though no implementation details were given, the author provided a description with citations so that others could find and replicate the treatment. The mentee teachers were rated, based on four classroom observations taken on different days, in the following areas (though no tests of reliability were reported): instructional management, rules and procedures, meeting student concerns, management of pupil

behavior, and student misbehavior. In addition, time-on-task of students was observed, and the frequency with which mentor and mentee discussed classroom management was measured, based on written reports of conferences between the teacher pairs (no additional information was provided about the quality of the conference report summaries).

One limitation of this study is that no numerical data were reported, only textual summaries of the findings and significance tests. Overall, the two treatment groups (in which the mentee and/or the mentor received the classroom management training) did not differ from each other on any of the measures; however, both of the treatment groups had more on-task behaviors in their classrooms and discussed management problems less often with their mentors than did beginning teachers in the control group. On the other classroom observation measures, there were no differences between the control and treatment groups.

Other limitations of the research include that the classroom management training took place mid-year, that one semester might not have been long enough for stronger effects to take place (especially since the treatment occurred mid-year), and that only one observer rated each teacher.

The **Schaffer, Stringfield, and Wolfe (1992)** evaluation used a pretest/posttest design on one group of beginning teachers who participated in an induction program.⁵ The following research questions, as stated by Schaffer et al., are of primary interest in this review: (1) To what extent did beginning teachers' behaviors change during a two-year structured induction program? (2) Which changes occurred during the first year, during the second year? Nineteen beginning teachers were selected from a group of 45 teachers who had volunteered to take part in

⁵ The authors also studied a group of experienced teachers (with at least five years of experience) who received research-based information about the effective use of class time, so differences in change between the group of beginning and experienced teachers could be compared; however, since this review is focused on the impact of induction efforts on beginning teachers, we are not focusing on the comparisons with experienced teachers.

the induction program. Selection was based on the entrance criteria for a master's program, including scores on the National Teacher Examination, undergraduate GPA, and interviews.

In the first year of the two-year induction program, the program faculty worked with schools and district personnel to design curriculum for the beginning teachers and to select and involve mentor teachers in that process. Beginning teachers received three hours of instruction per week, along with behavioral feedback and support focused on improving their skills in classroom organization, management, and instruction. Individual problem-solving sessions were also held. Each participant received results of his/her classroom observations to use in setting goals for behavioral changes. In the second year of the program, the emphasis shifted to alternative instructional models and material on higher-cognition questioning, matching methods to content and goals, peer coaching, and professional development growth plans. The program was based on documented research into classroom practices, was designed in line with what is known about adult learning and delivering effective professional development, and was targeted to improve classroom teaching.

In the first year of the evaluation, each beginning teacher was observed in the fall and spring for three one-hour sessions; in the second year, each beginning teacher was observed for four hours (two in the fall, two in the spring) following a standardized observation schedule. The authors reported interrater reliability for five of the six observers to be between 0.80 and 0.92. Three aggregate variables—academic statements, organizing statements, and behavior-related statements—were created from 51 observation measures. For each of these variables, standards from previous research were available to indicate the percentage of statements necessary to evidence effective teaching (e.g., at least 80 percent academic statements, less than 12 percent organizing statements, and less than 3 percent behavior-related statements).

Overall, the results showed positive changes on the part of beginning teachers, and the data presented included an adequate reporting of significance tests. During their first year of teaching, the participant teachers started below the effective percentage for academic statements, improving over the year to achieve an average of 80 percent at the end of the year. They began above the effective percentage of less than 12 percent organizing statements, and although they made a significant improvement (dropping from 22 percent to 15 percent), they finished the year still above the target value. There was no change in their behavior-related statements, which at 3.8 percent remained somewhat above the target value. Teachers who remained for a second year in the program showed a pattern for academic and organizing statements: they started the second year at a level better than their end point of the first year, declined somewhat over the year, but remained still improved over their end-of-the-first-year performance. In behavior-related statements, the teachers improved (by making fewer statements) over the second year and achieved the target goal of less than 3 percent.

One major limitation of this study is that there was no comparison group of beginning teachers, making it impossible to distinguish change in teaching behavior due to the treatment from change that occurs naturally or developmentally over the course of a beginning teacher's first two years of teaching. In addition, the findings are limited to beginning teachers who were interested and qualified to participate in a two-year master's program; the treatment may have been more intensive than many induction programs can provide, and, thus, a highly selective sample of beginning teachers not representative of the larger population of new teachers may have participated in this research. The work also could have been strengthened if it had used multiple outcome measures.

Ratsoy et al. (1987) evaluated Alberta's Initiation to Teaching Project by using a quasi-experimental design with some pretest/posttest comparisons and some posttest-only comparisons. From this large-scale evaluation, the portion relevant to this review is the Classroom Observation Study. Of particular interest are the following research questions addressed by the project: (1) Do first-year teachers who take part in an internship induction program have higher levels of teaching competencies at the end of the first year compared to the beginning? (2) Are there significant differences in the teaching competencies of teachers who go through the intern program compared to beginning teachers who do not? Alberta's Initiation to Teaching Project provided interns with a gradual transition to full-time teaching, along with professional assistance. Interns worked with a supervising teacher and taught an average of 50 percent time at the beginning of the year and 75 percent time at the end of the year; they also conducted classroom observations and attended workshops, conferences, parent interviews, and professional development activities. Implementation, duration, and structuring of the program varied across school districts. Given the flexibility built into the program, it would have been helpful to present summary statistics indicating the mean number of hours interns spent per week with the supervising teacher.

Participants in the study included 151 interns and 120 first-year teachers during the first year of the research (1985) and 92 interns, 51 first-year teachers, and 97 second-year teachers during the second year of the research (1986). Interns and beginning teachers were randomly selected for inclusion in the study from the pool of internship participants (volunteers) and the pool of first-year teachers in the province, respectively. In terms of comparability of these two groups of teachers, the authors reported that there were no differences in the universities they had

attended but that the first-year teachers who were not interns had higher GPAs and higher grades on the teaching practicum than those who participated in the internship program.

A classroom observation instrument was used to rate teachers on classroom competencies; teachers were evaluated by trained observers on 26 teaching strategies—e.g., rules and routines, listening, pace, smooth flow, clear information, and responsiveness—rated on a scale from 1 to 5 (where 1 is unacceptable and 5 is superior). Though the authors did not provide details about the observation instrument and its implementation in this publication, they indicated that similar measures have been widely tested and used, that adequate interrater reliability was achieved, and that the measures were not grade-level or subject specific.

Pre- and posttest scores (from observations in fall 1985 and fall 1986) were compared for interns and for first-year teachers. Results indicated that there were no statistically significant differences in the mean competencies of these two groups. On posttests in 1986, the beginning teachers who had been interns in 1985 had higher mean ratings on 22 of the 26 teaching strategies, compared with their pretests at the beginning of the internship year. Comparison group teachers also improved in the second year of the evaluation on 22 of the 26 strategies. This part of the study did not address whether one group of teachers changed more than another. In 1986, during the second year of teaching, former interns had significantly higher mean scores on 21 of the 26 strategies than teachers with no internship participation.

One limitation of this work is that the treatment did not involve full-time classroom responsibilities, so the results may not be generalizable to contexts in which teachers in induction programs are also full-time teachers, as is the case in most U.S. programs. Also, it was not made clear whether people who obtained a full-time job in the second year of the study were

systematically different from those who did not obtain a teaching job; hence, the value of the internship for those who did not obtain teaching positions is not known.

Gunter's (1985) evaluation of the Florida Beginning Teacher Program examined the impact of induction on teacher quality, as measured by beginning teachers' self-ratings of confidence in their teaching skills, beginning teachers' performance on the Summative Observation Instrument (a measure of teaching performance), and principals' recommendations about whether the beginning teacher should continue to be employed. The Summative Observation Instrument assessed the following domains: instructional organization and development, presentation of subject matter, verbal and nonverbal communication, and classroom management. According to the author, each measure had a history of use. (See section on retention research for more details about the study's sample, design, limitations, and findings about retention.) This induction program—which provided beginning teachers with a support team, a professional development plan, classroom observations with feedback, and assistance creating a portfolio—was found to have a limited impact on its participants in terms of teacher quality: program participants used significantly fewer teaching strategies identified to be ineffective by the researchers than the comparison group of beginning teachers. No statistically significant differences were found between the treatment and comparison groups in the average rating of self-confidence in teaching, principals' recommendations for continued employment, or the use of teaching strategies thought by the researchers to be effective. It should also be noted that no tests of reliability for the observation measures were reported.

Kilgore and Kozisek (1989) used primarily teacher self-reports and some observation data to study the impact of a college/university-based induction program on beginning teachers, with a focus on teachers' perceived knowledge base, administrators' perceptions of their skills,

teachers' expectations and realities concerning teaching, and the types of in-school support received by teachers as part of the program. Qualitative data were also collected, though these are not a focus in this review. The induction program studied was developed by Teachers College, University of Nebraska; Concordia Teachers College; and Doane College (all in Nebraska and part of a statewide consortium) and included a summer graduate program, first-year on-site visits, and regional seminars. As part of the program, participants also received assistance finding a teaching position and preparing for the job, a mentor teacher (in most cases, although no information was provided about mentor training), and on-site visits by college personnel, among other supports. The first-year teacher inductees also took part in two workshops: one helped new teachers develop skills in observation, testing, and test construction and helped socialize teachers and give them realistic expectations about the first year of teaching; the other workshop helped teachers in the planning and production of materials such as a calendar, bulletin boards, lesson plans for the first two weeks of school, a curriculum outline for the year, tests and quizzes, a discipline plan, transparencies, and plans for developing and reporting grades.

The authors implemented a quasi-experimental research design, with pre- and posttest measures for a treatment group of 22 first-year teachers who chose to participate in the induction program and a comparison group of 18 first-year teachers who chose not to participate. Results from the research did not indicate an impact of the induction program on beginning teachers. Both groups of teachers rated their teaching competencies as high at both the beginning and end of the first year of teaching (with no statistically significant differences). The teaching competencies surveyed included, among others, classroom discipline and management, evaluation of students, classroom organization, time management, teaching methods, use of

curricular materials, and communication with staff, students, and parents. There were also no differences between teacher groups in principals' ratings of their teaching performance, and it was found that first-year teachers and principals had different perceptions of the elements necessary for teaching effectiveness. Both groups of teachers brought expectations to the first year of teaching that were not met. Similar support services were offered to both groups, though not at a high level for either group.

It is difficult to assess the generalizability of the findings from Kilgore and Kozisek's study since the entry requirements to participate in the induction program were not made clear. The induction program sounds like a selective one, so it may be that the participants in both groups were not representative of all beginning teachers, and not enough description was given about the treatment and comparison groups to address this concern. Also potentially confounding the research, the control group of beginning teachers had chosen *not* to participate in the induction program (though they had been accepted), and no information was provided to assess whether this group of teachers was different in other ways from the experimental group that did choose to participate.

The **New York City Board of Education (1993)**, in its evaluation of the Mentor Teacher Internship Program, examined the impact of induction on beginning teacher effectiveness and socialization into the school community, as measured on self-report questionnaires. (See section on retention research for more details about the study's sample, design, limitations, and findings about retention.) In the NYC program, interns (with trained mentors to coach them) reported a greater improvement in their comfort level with disciplining/managing the classroom, writing lesson plans, working with fellow faculty, and confidence in their teaching, compared with teachers in the comparison group (statistically significant differences reported in these areas).

Also, longer mentoring (i.e., being assigned mentors at the beginning of the year versus mid-year) was associated with higher self-reports of teaching abilities.

Henry (1988a), also referenced in 1988b and 1990) described an evaluation of Project CREDIT (Certification Renewal Experiences Designed to Improve Teaching), a cooperative program between Indiana State University and 10 school districts in west-central Indiana. Project CREDIT sought to link university and public schools in teacher education, improve the teaching skills and reduce the problems of first-year teachers, reduce teacher burnout and attrition, and reward superior teaching. To accomplish these goals, beginning teachers in the program were assigned a mentor to be involved and help them on a daily basis, to provide role modeling, and to give formative growth experiences. Other forms of support were regular visits and observations by university faculty skilled in the supervision of field experiences; university affiliates also provided consultant services in classroom management, testing and evaluation, human relations, and the use of computers in the classroom. Additionally, beginning teacher participants in Project CREDIT attended monthly seminars with other first-year teacher participants outside of their schools. These seminars allowed for sharing, as well as focused conversations on specific topics related to teaching. Mentors for Project CREDIT interns were selected by building principals and the project director, and were required to be effective teachers with at least five years of teaching experience, to have experience in the building where the intern was teaching, and to be willing to devote extra time to serve as mentors, among other criteria. They were also paid a stipend.

Only the part of this evaluation that focused on teacher quality compared program participants with a comparison group of beginning teachers. Though the results were reported as statistically significant, almost no numerical data were presented. A comparison of 20

participant teachers' and an unspecified number of nonparticipant teachers' attitudes and perceptions about teaching revealed significant differences as follows. Interns' reports indicated a gain in the use of mastery learning and mastery learning theory, increased motivation to understand and use higher-order questions, increased motivation to teach critical-thinking skills, increased awareness of state and local curriculum guidelines, and enhanced abilities to communicate with parents and the public. In addition, teachers who participated in Project CREDIT had what the researcher described as significantly "healthier" attitudes and perceptions about teaching than did the comparison group—i.e., beginning teachers in the induction program with multiple forms of support were better able to cope in 88 of 98 surveyed areas. The author concluded that the induction program seemed to "intercept" declines in beginning teachers' attitudes toward teaching.

In addition to not presenting descriptive data or specific parameter estimates from tests of statistical significance, this evaluation lacks a description of the beginning teachers in both the treatment and comparison groups. All we know is that they are first-year teachers in west-central Indiana and that the treatment group is composed of the first 20 teachers employed in the 10 districts participating in the collaborative university-district teacher induction program. This lack of information makes it difficult to assess the comparability of the treatment and comparison groups, as well as the generalizability of the results. It is not clear whether the program participants, as earliest hires, were systematically different from comparison group teachers. Another limitation of the work is that it is not clear whether the Project CREDIT participants and nonparticipants were teaching in the same schools, which could introduce contamination into the treatment.

In their two-year study of Toronto's Peer Support Pilot Project, **Cheng and Brown (1992)** considered the impact of the comprehensive induction program—including mentoring, workshops, professional development plans, and release time for professional dialogue—on beginning teachers' perceived competencies and self-reported areas of difficulty in teaching, as surveyed at the end of the school year. (See section on retention research for more details about the study's sample, design, limitations, and findings about retention.) In the first year of the evaluation, pilot program teachers were more likely than comparison group teachers to report overall positive experiences as a new teacher; in the second year of the study, pilot program teachers were slightly more likely than the comparison group to report overall positive experiences. Pilot program teachers also found teaching less difficult at the end of the school year than they did at the beginning, whereas comparison group teachers found teaching as difficult or even more so. In terms of areas of difficulty in teaching, both groups of beginning teachers found organizational strategies the most difficult during the first year of the evaluation and instructional strategies the most difficult in the second year of the study; in both years, teachers found emotional support the least difficult area of teaching. No tests of statistical significance were reported for these findings.

The one study in this review that considered student achievement is the **Dana Center's (2002)** evaluation of the Texas Beginning Educator Support System (TxBESS). Beginning teachers who participate in the TxBESS program receive instructional and mentor support as well as a formative assessment during the first year of teaching. (See section on retention research for more details about the study's sample, design, limitations, and findings about retention.) In the second year of the evaluation, the authors examined student achievement on the math and reading parts of the Texas Assessment of Academic Skills (TAAS). They gathered

year-one and year-two achievement test scores for students whose teachers were beginning teachers in year two. Although they attempted to obtain data from each district in the study for beginning teachers who had participated in the program and a comparison group who had not participated, this was not possible in most of the districts. As a result, the two groups of students may not have been comparable. Controlling for students' first-year achievement score, ethnicity, grade level, and socioeconomic status, the authors found no statistically significant relationship between students' second-year achievement and teachers' participation in the program. This finding held for reading and math.

Conclusions

Although a large number of written resources on teacher induction exist, there are very few rigorous studies that have investigated the impact of induction on teacher quality and retention. Within the small body of research that has been conducted with an experimental or quasi-experimental design, studies of induction have been weak for a number of reasons: the scholarship lacks adequate definitions of the constructs of induction, retention, and teacher quality; the researchers often rely solely on self-report to measure the outcomes of interest; and many of the studies use only one outcome measure and do not always align the outcome assessed with the treatment(s) teachers receive. In addition, there are many potentially confounding factors in the research we reviewed—e.g., a lack of comparability between groups of teachers, contamination of the treatment groups by the presence of comparison groups in the same school, problematic timing of the treatment, weak control of the measurements gathered, and poor attention paid to the attrition of research participants. Taken together, although this research includes some positive findings, the studies are not strong enough for us to conclude that induction works—that it improves teacher retention or effectiveness (measured in terms of

student achievement or otherwise). This is not to say that induction does *not* work, only that the limitations of the current literature prevent us from drawing conclusions about the impact of induction on beginning teachers.

At the outset of conducting this literature review, one of our goals was to determine which types or combinations of induction activities have the greatest impact on the retention and effectiveness of beginning teachers. However, the dearth of high-quality experimental and quasi-experimental research in this area precludes us from pinpointing the most effective induction practices. Mentoring was a feature in all but one of the induction evaluations we reviewed, though there was not conclusive evidence overall about the impact of this commonly used method of inducting new teachers into the profession. It may be that induction programs that include, in addition to mentoring, other activities such as structured peer coaching, classroom observations with feedback, workshops, professional development plans, portfolio use, and so on, have the greatest impact; *or* it may be that mentor-based induction programs do not necessarily have to include a wide range of activities but are most useful when they focus on instructional practices with built-in opportunities for classroom observation and feedback. Existing studies on induction, though, do not answer the question of which components of induction have the strongest potential to improve the effectiveness and retention of beginning teachers.

References

- Brown, J. G., & Wambach, C. (1987, February). *Using mentors to increase new teacher retention: The Mentor Teacher Induction Project*. Paper presented at the annual meeting of the American Association of Colleges for Teacher Education, Arlington, VA.
- Cheng, M., & Brown, R. S. (1992). *A two-year evaluation of the Peer Support Pilot Project: 1990-1992*. Toronto, Ontario, Canada: Toronto Board of Education, Research Department.
- Charles A. Dana Center. (2002). *Texas Beginning Educator Support System: Evaluation report for year three, 2001-02*. Austin: University of Texas, College of Natural Sciences.
- Fideler, E. F., & Haselkorn, D. (1999). *Learning the ropes: Urban teacher induction programs and practices in the United States*. Belmont, MA: Recruiting New Teachers, Inc.
- Gold, M. (1987). *Retired teachers as consultants to new teachers: A new inservice teacher training model. Final report. Case 09-87*. New York: City University of New York, New York Institute for Research and Development in Occupational Education; Washington, DC: American Association of State Colleges and Universities.
- Gunter, M. N. (1985). *A comparative study of participants and nonparticipants in the Florida Beginning Teacher Program*. Fort Myers, FL: Florida Educational Research and Development Council.
- Henry, M. A. (1990). Project CREDIT: Case study of a cooperative program in teacher education. *Contemporary Education*, 61(3), 152-156.
- Henry, M. A. (1988a). *Project CREDIT. Certification Renewal Experiences Designed to Improve Teaching. A cooperative multiple-support program for first-year teachers between Indiana State University and ten public school districts*. Terre Haute: Indiana State University, Department of Secondary Education.
- Henry, M. A. (1988b). Multiple support: A successful model for inducting first-year teachers. *The Teacher Educator*, 24(2), 7-12.
- Kilgore, A. M., & Kozisek, J. A. (1989). The effects of a planned induction program on first-year teachers: A research report. In J. Reinhartz (Ed.), *Teacher induction. NEA aspects of learning*. Washington, DC: National Education Association.
- Klug, B. J., & Salzman, S. A. (1991). Formal induction vs. informal mentoring: Comparative effects and outcomes. *Teaching & Teacher Education*, 7(3), 241-251.
- New York City Board of Education, Office of Research, Evaluation, and Assessment. (1993). *Mentor Teacher Internship Program, 1992-93. OERA report*. Brooklyn, NY: Author.
- Ratsoy, E., et al. (1987). *Evaluation of the Initiation to Teaching Project. Final report*. Edmonton, Alberta, Canada: Alberta Department of Education.

- 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(3), 181-192.
- Smith, T. M., & Ingersoll, R. M. (2003). *Reducing teacher turnover: What are the components of effective induction?* Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.
- Stallion, B. K. (1988, February). *Classroom management intervention: The effects of training and mentoring on the inductee teacher's behavior.* Paper presented at the annual meeting of the Association of Teacher Educators, San Diego, CA.
- Stallion, B. K., & Zimpher, N. L. (1991). Classroom management intervention: The effects of training and mentoring on the inductee teacher's behavior. *Action in Teacher Education*, 13(1), 42-50.

Appendix A: Summaries of Literature Reviewed

Reference	Research Questions or Objectives	Design	Sample Size and Characteristics	Data Collection Method(s)
Brown & Wambach (1987)	How will an induction program with a support system that emphasizes (1) contact with a mentor, (2) involvement of mentor teachers in seminars, and (3) opportunities for both formal and informal professional support affect beginning teachers (both student teachers and first-year teachers)?	Quasi-experimental design with a treatment group and a comparison group for each of two populations: student teachers and first-year teachers. Change in student teachers' attitudes was measured by pretest and posttest attitudes; all other measures were posttest only.	Sample size and characteristics of the sample are not reported.	Post-program surveys.
Cheng & Brown (1992)	To ascertain the degree of success of the Toronto Board's Peer Support Project as a formal induction program.	Quasi-experimental design with a treatment and comparison group. Some analyses with pretest/posttest measures, some with posttest only measures.	1990-91: 25 pairs of teachers (beginning teacher and mentor, at the same school and grade level or program) participating in the Toronto Board's Peer Support Project. 1991-92: 38 pairs of teachers. Comparison group of 30 beginning teachers in 90-91, and comparison group of an unspecified number of beginning teachers in 91-92.	About beginning teachers: monthly journals (1990-91 only), questionnaires (three times in 1990-91, Sept., Jan., June; and two times in 1991-92, Jan., June), and a focus group to discuss the program (1991-92).
Charles A. Dana Center (2002)	What are the characteristics of the 1999, 2000, and 2001 cohorts of TxBESS participants—beginning teachers, mentors, and support team members? What are the retention rates of beginning teachers in Texas? What is the relationship between student performance and teacher characteristics? (Other evaluations were made as part of the larger study.)	Quasi-experimental design with treatment and comparison group and posttest measures.	1999-2000: 998 participants; 2000-2001: 2,059 participants; 2001-2002: 3,058 participants. TxBESS teachers are compared with non-TxBESS beginning teachers in the state.	TxBESS and non-TxBESS beginning teachers are compared using the TEA-PEIMS database to track retention over a three-year period. Surveys administered to all TxBESS teachers, mentors, principals, and teacher preparation instructors. Student achievement data from the state database on the six largest districts in the state were compared in the aggregate to identify growth.
Gold (1987)	To evaluate a mentor program with the goals of improving teaching and reducing attrition of first-year teachers.	Experimental design with posttest of treatment and control group. Mentors were assigned to a school, and then principals assigned the mentors at random three new teachers as mentees. Three other new teachers in the school served in the control group.	Sample characteristics are not clear except that teachers worked in schools with high levels of teacher turnover. Report indicates that the program has about 100 mentors and 300 mentees per year. Retention evaluation was based on only 160 mentored teachers and 113 comparison teachers.	Teachers completed surveys twice during the school year to measure the support received by mentors (or by supervisors, in the case of the control group teachers). Method of attrition measurement was not specified.

Reference	Research Questions or Objectives	Design	Sample Size and Characteristics	Data Collection Method(s)
Gunter (1985)	Do inexperienced teachers in their first year of teaching who participate in the Beginning Teacher Program differ in attitude, behavior, and/or competence from inexperienced, first-year teachers who do not participate in the program?	Quasi-experimental. A treatment group and a comparison group matched for age, race, sex, grade level taught, subject area taught, degree, and university (in-state or out-of-state). Analysis based on posttest measures.	77 matched pairs; one teacher participated in the program, the other did not. Program participation was determined by the date when teachers applied for certification: those who applied before a new law went into effect were not required to participate; those who applied after were subject to the new requirement.	Questionnaires administered at the end of the program and observations made at the end of the school year.
Henry (1988a)	To measure the extent to which Project CREDIT personnel accomplished the stated objectives of the induction program, and to measure changes in intern attitudes and perceptions during the course of the program. Stated objectives of Project CREDIT include: linking universities and public schools in teacher education, improving the teaching skills and reducing the problems of beginning teachers, reducing teacher burnout and attrition, and rewarding superior teachers.	Quasi-experimental. Treatment and control group with pre- and post-program measures of teacher effectiveness.	Treatment group: 20 first-year teachers in west-central Indiana who were participating in Project CREDIT; this group was composed of the first 20 teachers who were employed in the 10 participating school districts for the 1986-87 school year. Control group: unspecified number of first-year teachers in west-central Indiana who were not participating in the program. An effort was made to match mentors and new teachers in terms of subject area, though the overriding criterion was whether the two were expected to work well together.	Pre- and post-program self-reports of attitudes and perceptions about teaching and their teaching effectiveness.
Kilgore & Kozisek (1989)	To evaluate the effects of a college/university-based induction program on beginning teachers, in terms of their perceived knowledge base, their expectations and realities concerning teaching, administrators' perceptions of their skills, and reports about the types of in-school support they received.	Quasi-experimental. Treatment and control groups with pretest and posttest measures.	Experimental group: 22 teachers who chose to participate in the induction program. Control group: 18 teachers who chose not to participate in the program.	Quantitative data: self-reports of perceived level of knowledge and competency on selected teaching behaviors (at the beginning and end of the year), principal ratings of the same behaviors, and teachers' questionnaire responses about their expectations of working conditions (prior to the school year) and the realities they experienced (at the end of the year). Qualitative data: weekly logs kept by teachers, on-site observations by researchers, notes taken from regional seminars held for the teachers.
Klug & Salzman (1991)	To compare the impact of two induction approaches on teacher competencies, attitudes, and morale, among a group of first- and second-year teachers, randomly assigned to each group.	Experimental. Participants assigned at random to one of two treatments (formal induction with a mentor and informal mentor-based induction), tested with pre- and posttest.	26 first- or second-year teachers in three rural or small city school districts in southwest Idaho. All were assigned a mentor, though it is not clear whether mentors were matched with teachers on any characteristics.	Videotaped classroom observations, rated by researchers using two of five scales of the Teacher Performance Assessment Instrument (TPAI) to assess teacher competencies. Purdue Teacher Opinionnaire, on which teachers responded to 145 statements (1-4 scale) about morale and attitudes. Structured qualitative interviews that took place at the end of the school year.
New York City Board of Education (1993)	Does participation in the NYC Mentor Teacher Internship Program improve teaching effectiveness and socialization into the school community, improve instruction for students, and increase teacher retention?	Quasi-experimental design with a treatment and control group measured with posttests.	864 mentors, representative of all the NYC school districts; 1,264 interns, newly hired in NYC public schools, without certification; 100 control group teachers, newly hired and not yet matched with a mentor.	Retrospective surveys completed by mentors, interns, control group teachers (and others not focused on in this review) in May. Mentor logs were also examined.

Reference	Research Questions or Objectives	Design	Sample Size and Characteristics	Data Collection Method(s)
Ratsoy et al. (1987)	(1) Do first-year teachers who take part in an internship induction program have significantly higher levels of teaching competencies at the end of the first year of teaching compared with the beginning of year? (2) Are there significant differences in the teaching competencies of teachers who go through the intern program compared with beginning teachers who do not go through the program?	Quasi-experimental using the following designs: (1) One-group pretest/posttest of classroom competencies via trained observers. Interns were observed in fall 1985 during the internship year and again in fall 1986 during the first year of full-time teaching. First-year teachers who did not participate in the intern program were observed in fall 1985 and again in fall 1986 as second-year teachers. This part of the study did not include comparisons between interns and first-year teachers. (2) Posttest only, with comparison group in 1986; teachers who had gone through the internship in 1985 and then obtained teaching jobs were compared with second-year teachers who had not participated in the internship. In 1986, a sample of first-year teachers with no previous teaching experience was added to the control group (i.e., the control group that year contained first- and second-year teachers with no internship experience).	1985: 151 interns, 120 first-year teachers. 1986: 92 interns, 97 second-year teachers, and added 51 first-year teachers. Participants in the study were randomly selected from the pool of first-year teachers in the province and the pool of internship participants (who volunteered). The authors reported that there were no overall differences in the universities attended, but that the first-year teachers had overall higher GPAs and higher grades on teaching practicum than did those in the internship program.	A classroom observation instrument was used to rate teachers on 26 classroom competencies (e.g., rules and routines, listening, pace, smooth flow, clear information, and responsiveness), scored on a scale from 1 to 5. Full description of the measure is included in a separate technical report, but the authors indicated that similar measures have been widely tested and used.
Schaffer, Stringfield, & Wolfe (1992)	(1) To what extent did beginning teachers' behaviors change during a two-year structured induction program? (2) Which changes occurred during the first year, during the second year? (3) How did changes in inductees' teaching compare with those of highly experienced educators who were involved in a similar though less intensive experience focused on the effective use of classroom time?	Quasi-experimental design having pretest and posttest measures on two groups of teachers: beginning teachers and experienced teachers, each of which received some training. Beginning teachers were followed into their second year of teaching with pre- and posttest measures that year as well.	19 beginning teachers, 50% of whom were volunteers, and 12 experienced teachers with at least five years of experience, all of whom volunteered to take part in training on the use of class time. The beginning teachers were selected for the master's program on the basis of scores on the NTE, undergraduate GPA, and interviews. Participants had slightly higher GPAs than those who were not selected.	Data were gathered through classroom observations using a low-inference rating system that had a long history of use and development in classroom research. In the first year, each teacher was observed in the fall and spring for three sessions, one hour each session; in the second year, the beginning teachers were observed for four hours (two hours in the fall and two in the spring).
Stallion (1988)	To assess the effects of classroom management training, embedded in a teacher induction program, on beginning (and mentor) teachers, and to assess the effects of mentoring relationships on the beginning teachers' classroom management behaviors.	Experimental design with posttest measures. Mentor/mentee pairs assigned at random to one of three groups, stratified by grade level and subject. Group 1: mentor and mentee participate in classroom management training, Group 2: only mentor participates, Group 3: neither participates.	35 mentee/mentor pairs (matched by their principals on grade level or subject) who were involved in a school/university-based induction program. No information on number of pairs per group. Mentees were teachers new to their classrooms and included teachers returning from leave, teachers teaching at a new grade level or in a new subject matter, and teachers new to their school or district.	Classroom observations and reviews of written reports of mentor/mentee conferences.

Reference	Treatment(s) and Independent Variables	Quality of Treatment Definition	Outcome(s) and Dependent Variables	Quality of Outcome(s) Definition(s)
Brown & Wambach (1987)	STUDENT TEACHERS: assigned to a mentor teacher for a seven-week student-teaching experience; in addition to providing students with teaching experience, mentors attended three of the seven weekly seminars held for the student teachers. FIRST-YEAR TEACHERS: matched to a mentor of similar teaching assignment for one year of support including regular ongoing contact to discuss problems and concerns, visitation of the beginning teacher to the mentor's class and of the mentor to the beginning teacher's class, involvement in program seminars; mentee also received monthly support sessions with professional counselors and university faculty. BOTH GROUPS: worked with mentors who were experienced teachers and applied or were recommended for the role; the mentors received a small stipend for their time.	Weak. Insufficient information about the topics and formats of the seminars. No information on the training of mentors or what they did in working with teachers.	(1) Change in student teachers' attitudes toward teaching, students, and the school; (2) student and first-year teachers' intent to stay in teaching; and (3) student and first-year teachers' judgment of their teaching experience during the program, and student and first-year teachers' judgment of the success of the program.	Weak. Student teacher attitudes were measured using two attitude scales; however, all other measures were made based on single items from a survey.
Cheng & Brown (1992)	Orientation meeting (both years); meetings in October and May to share common issues and concerns (1990-91); series of three half-day workshops in November, February, and April (1991-92) on themes identified by participants as important for the professional development of new teachers (e.g., whole language learning); and five days of release time for professional dialogue and sharing (both years), including the creation of personalized professional development plans.	Adequate.	Experiences as a first-year teacher (e.g., decision to become a teacher, choosing teaching as a career again, staying in the teaching profession, areas of difficulty in teaching) and perceived competence.	Adequate self-report measures.
Charles A. Dana Center (2002)	Teachers participating in the TxBESS program are provided with instructional and mentor support during their first year of teaching. The program offers comprehensive support, training, and formative assessment to assist beginning teachers in Texas public schools. Student achievement data were obtained from the statewide database and statistically controlled for ethnicity, grade level, and SES.	A comprehensive program of support, training, and formative assessment took place according to performance standards during beginning teachers' first year of teaching (for both the teacher and the support team).	Retention and student achievement as measured by the mathematics and reading portions of the TAAS.	Used an existing statewide database for the retention comparison and subscales from the statewide assessment TAAS, but the surveys were sent only to participants.
Gold (1987)	Retired teacher with training as mentor, for 66 hours per school year.	Good. Information was provided to describe the recruitment, selection, training, and responsibilities of mentors. In addition, a survey of teachers with and without mentors documents differences in supports received during the study year.	Attrition from teaching.	Method of measurement and definition of attrition (e.g., leaving profession vs. leaving school) are not provided. Also, poor survey response rate.

Reference	Treatment(s) and Independent Variables	Quality of Treatment Definition	Outcome(s) and Dependent Variables	Quality of Outcome(s) Definition(s)
Gunter (1985)	PROGRAM PARTICIPANTS were assigned a support team (administrator, peer teacher, and another educator) who created a professional development plan for them. Goals were linked to a performance measurement system that was research-based. Each team member was required to make at least one observation of the participant, including a pre-observation conference to select the setting for the observation and a post-observation conference to provide feedback. The administrator also made two observations during the year and maintained a portfolio containing documentation of observations and conferences as well as examples of teacher-designed materials. Upon completion of the program, the administrator evaluated the participant and recommended successful or unsuccessful completion of the program. In-service training was provided to principals before using the observation instrument. Program participants' training was the responsibility of their district, ranging from six to 20 hours, and covering the Florida Performance Measurement System. NONPARTICIPANT TEACHERS completed their first year of teaching without an identified support system or training, but had regular district evaluations two to three times/year.	Treatment seems minimal in providing support for teachers since the amount and quality of support/training depended on how the program was implemented locally. There was not a formal planning and assessment component to the research.	Teacher absentee rate, score on a job satisfaction questionnaire, self-rating on confidence level in performing teaching skills, statement of intent to remain in the teaching profession, recommendation by principal for continued employment, performance on the Summative Observation Instrument (a measure of teaching performance).	Multiple variables were examined, each with a measure that had some history of use.
Henry (1988a)	Beginning teachers in the program were assigned a mentor to be involved and help them on a daily basis, to provide role modeling, and to give formative growth experiences. Other forms of support were regular visits and observations by university faculty skilled in the supervision of field experiences; university affiliates also provided consultant services in classroom management, testing and evaluation, human relations, and the use of computers in the classroom. Additionally, beginning teacher participants in Project CREDIT attended monthly seminars with other first-year teacher participants outside of their schools. These seminars allowed for sharing, as well as focused conversations on specific topics related to teaching. Mentors for Project CREDIT interns were selected by building principals and the project director, and were required to be effective teachers with at least five years of teaching experience, to have experience in the building where the intern was teaching, and to be willing to devote extra time to serve as mentors, among other criteria. They were also paid a stipend of \$2,000.	Treatment provided multiple forms of support to first-year teachers—mentors and input from experienced teachers as well as university faculty, and monthly seminars to interact with other new teachers. A good description of mentor selection criteria was provided, as well as the areas of support they were supposed to focus on in their work with the new teachers; however, there was no discussion about whether or how the mentors were trained. There also seemed to be no formal assessment component to the program.	Attitudes and perceptions of first-year teachers about teaching and about their teaching effectiveness.	No specifics are given about the questionnaires used to gather information from beginning teachers about their attitudes and perceptions about teaching and about their teaching effectiveness.
Kilgore & Kozisek (1989)	First-year teacher support program developed by Teachers College, University of Nebraska, Concordia Teachers College, and Doane College (all in Nebraska and part of a statewide consortium). Includes summer graduate program, first-year on-site visitation program, and regional seminars. Participants received help finding a teaching position and preparing for the job, a mentor teacher (in most cases), and on-site visits by college personnel, among other things. They also participated in two workshops: one helped new teachers develop skills in observation, testing, and test construction and helped socialize teachers and give them realistic expectations about the first year of teaching; the other workshop helped teachers in the planning and production of materials such as a calendar, bulletin boards, lesson plans for the first two weeks of school, a curriculum outline for the year, tests and quizzes, a discipline plan, transparencies, and plans for developing and reporting grades.	A good description of the program was provided, though there is no discussion of the training of mentors.	Perceived knowledge base held by first-year teachers, administrator perceptions of first-year teachers' performances on knowledge base criteria, and first-year teachers' expectations and realities of teaching.	Adequate self-report measures and principal ratings of teacher competencies.
Klug & Salzman (1991)	Group 1 received classroom observations by mentor team two hours per month during the first semester; semi-monthly observations during the second semester (debriefings not described); two team meetings per semester; overall, mentors spent about 72 hours with new teacher over the year. Group 2 underwent informal mentor induction (a buddy system) with no structured number of hours and no guidelines about the content of the mentoring relationship; also, mentors did not receive any special training.	The treatment provided to Group 1 was clear. The comparison treatment provided to Group 2 was intentionally unstructured, and mentors spent an average of 22 hours with the first-year teachers and 12 hours with the second-year teachers over the course of the year.	Observer ratings on the TPAI; teacher responses to questionnaire; qualitative data from end-of-year interview	Very thorough description of validity and reliability of the TPAI and the questionnaire.

Reference	Treatment(s) and Independent Variables	Quality of Treatment Definition	Outcome(s) and Dependent Variables	Quality of Outcome(s) Definition(s)
New York City Board of Education (1993)	Mentors received training and release time and were matched with newly hired uncertified teachers, interactions included peer coaching to improve teaching effectiveness, the opportunity to take a graduate-level course provided by the United Federation of Teachers, and other ongoing professional development activities over the school year. The graduate courses available to teachers focused on teaching techniques and classroom management, specific content areas, or content or technique combined with a focus on specific student populations.	A good description of the program was provided so that the research could be replicated; however, all levels of the independent variables were unplanned and post hoc.	Improved teaching effectiveness and socialization into the school as measured by items on the survey instruments. Retention as measured by self-report on how long teachers planned to stay in the profession.	Defined by questions on the survey instrument, which was not shown, and no psychometric qualities of the instrument were reported.
Ratsoy et al. (1987)	The internship program intended to implement a gradual transition to full-time teaching, along with professional assistance. Interns worked with a supervising teacher and taught an average of 50% time at beginning of year, 75% time at the end; they also attended workshops, conferences, parent interviews, and professional development activities, and conducted classroom observations. Programs varied across school districts, and there was flexibility across sites in terms of the structure of the internship.	Given the flexibility built into the program, it would have been helpful to have some summary statistics on the mean number of hours spent with supervisory teacher per week, etc. It also would have been helpful to know the variability of key characteristics of the treatment.	Scores on a range from 1 to 5 on 26 teaching strategies (1 = unacceptable, 5 = superior) of the Classroom Observation Record.	Details about observation techniques are not available in this publication, though there is an adequate description of reliability and validity. Interrater reliability was tested and reported as adequate. The authors also tested to confirm that the measures of classroom competencies were not grade-level specific or subject specific.
Schaffer, Stringfield, & Wolfe (1992)	The main independent variable of interest was an intensive two-year induction program. Also of interest was the difference between changes in the teaching behavior of beginning vs. experienced teachers when both receive research-based information about effective use of class time. In the first year of the two-year induction program, beginning teachers received three hours of instruction, behavioral feedback and support (from clinical supervision) focused on improving skills in classroom organization, management, and instruction. Individual problem-solving sessions were also held. Each participant received results of his/her classroom observations to use in setting goals for behavioral changes. In year 2, the emphasis shifted to alternative instructional models and material on higher-cognition questioning, matching methods to content and goals, peer coaching, and professional development growth plans.	The induction treatment was based on sound research into classroom practices, was designed in line with what is known about delivering effective professional development and adult learning, and was targeted to improve classroom teaching.	The primary dependent variables were three aggregate variables created from 51 variables in the observation system: academic statements, organizing statements, and behavior-related statements. For each, standards from previous research were available to indicate the percentage of statements that were associated with effective teaching (e.g., at least 80% academic, < 12% organizing, and <3% behavioral).	Outcomes were well linked to the treatment, and the measure was a strong observational system with high interrater reliability. The study would have been stronger with more than one outcome measure.
Stallion (1988)	Classroom management intervention training took place mid-year. Workshop was designed on basis of research into classrooms to prevent management problems. Group 1: training for mentor and beginning teacher, Group 2: training for mentor only, Group 3: neither teacher receives intervention.	Instructional treatment based on solid research foundation; description gives cites so others could find and replicate it. No implementation evidence, however.	Teachers were rated on instructional management, rules and procedures, meeting student concerns, management of pupil behavior, and student misbehavior. Time on task of students also was observed. Finally, the frequency with which the mentor and mentee discussed classroom management concerns was measured by examining conference report forms.	Strengths: multiple measures of classroom management taken on four different days; examination of the generalizability of each observational measure. Weaknesses: only one observer, observer possibly not blinded to the experimental groups of the teachers, no information on the quality of the conference report forms.

Reference	Type of Analysis	Results	Adequate Reporting of Statistical Tests	Accurate Estimates of Effect Sizes
Brown & Wambach (1987)	Conducted t-tests to test for differences between groups.	FOR STUDENT TEACHERS: No significant differences in intent to stay in teaching or the extent to which they judged their student-teaching experience to be a success. For one measure of attitude there were no pre-to-post changes for either group; for the other measure the mean score of the control group decreased significantly from pre to post, while that of the program group remained unchanged. FOR FIRST-YEAR TEACHERS: There were no differences in reported intent to stay in teaching or in judgment of the success of their first year of teaching. FOR BOTH GROUPS: All teachers who participated in the program judged it to be a success.	Yes, statistical tests were conducted, though the test statistics were not presented except in one case.	Cannot compute from data provided.
Cheng & Brown (1992)	Percentages regarding teachers' experiences in their first year of teaching.	In 1990-91, pilot program teachers were more likely to report positive experiences as new teachers; all pilot program teachers felt they made the right career choice, and almost all said they would make the same decision again, compared with lower rates for the comparison group; more pilot project teachers (76%) expected to still be in teaching five years later than the comparison group (60%). In 1991-92, pilot program teachers were slightly more likely to report positive experiences as new teachers; almost all teachers in both groups thought they made the right career choice and would make the same decision again; slightly more teachers in the pilot project (97%) than in the comparison group (91%) expected to remain in teaching two years later. Pilot program teachers found teaching less difficult at the end of the year; the comparison group found teaching as difficult or even more so. Both groups found organizational strategies the most difficult in 1990-91, instructional strategies the most difficult in 1991-92, and emotional support the least difficult in both years; pilot program teachers perceived a reduction in most areas of difficulty, while comparison teachers perceived a minimal reduction or increase, in both years.	No tests of statistical significance.	None reported, but effect sizes could be computed from data provided.
Charles A. Dana Center (2002)	Frequencies and percentages regarding retention, and regression analyses of student achievement data.	Improved retention overall, especially for minority groups and high school teachers. No significant difference in math or reading scores for students of TxBESS and non-TxBESS teachers.	Significance tests were reported for the achievement data analyses but not the retention measures.	None reported, but effect sizes could be computed from data provided for the retention portion of the study.
Gold (1987)	Simple percentages reported for attrition, the dependent variable; for the measures of support, chi-square tests were conducted to test for differences in the two groups' responses.	Very low attrition rates (1.88% in mentored group, 3.54% in control group in the same school, 4.54% for teachers in similar schools with no mentors — no tests of statistical significance conducted. Survey responses indicated that mentors provided more support (to mentees) than school administrators (to control group teachers) on 40 of 43 activities.	None with regard to teacher attrition.	None reported, but effect sizes could be computed from data provided.

Reference	Type of Analysis	Results	Adequate Reporting of Statistical Tests	Accurate Estimates of Effect Sizes
Gunter (1985)	Paired t-tests conducted for the data that could be considered to be continuous; significance tests conducted for the dichotomous data.	AVG. NUMBER OF DAYS ABSENT: Differences in the expected direction though not quite statistically significant ($p=0.054$). AVG. JOB SATISFACTION: No difference. AVG. RATING OF SELF-CONFIDENCE IN TEACHING COMPETENCIES: No difference. PERCENT PLANNING TO REMAIN IN TEACHING: No difference. PERCENT PLANNING TO REMAIN IN SAME POSITION: Program participants were significantly less likely to plan to remain in same position. PRINCIPAL RECOMMENDATIONS FOR CONTINUED EMPLOYMENT: No difference. OBSERVED NUMBER OF EFFECTIVE TEACHING STRATEGIES: No difference. OBSERVED NUMBER OF INEFFECTIVE TEACHING STRATEGIES: Program participants used significantly fewer strategies considered ineffective.	Yes, tests of statistical significance were conducted and reported.	None reported, but effect sizes could be computed from data provided.
Henry (1988a)	Unclear from the write-up.	Significant differences: interns experienced a gain in the use of mastery learning and mastery learning theory, increased motivation to understand and use higher-order questions, increased motivation to teach critical-thinking skills, increased awareness of state and local curriculum guidelines, and enhanced abilities to communicate with parents and the public. In addition, teachers who participated in Project CREDIT had significantly "healthier" attitudes and perceptions about teaching than did the control group—i.e., beginning teachers in the induction program were better able to cope with 88 of 98 surveyed variables about attitudes or teaching perceptions. The author concluded that the induction program seemed to "intercept" declines in beginning teachers' attitudes toward teaching.	Almost no data were provided; author notes "significant" findings but does not provide any description of analyses nor parameter estimates.	Cannot compute from data provided.
Kilgore & Kozisek (1989)	Percentages, means (with standard deviations), t-tests.	Both groups of teachers rated their teaching competencies as high at the beginning and end of their first year of teaching (no statistically significant differences). There were also no differences between groups in principals' ratings of their teaching performance, and it was found that first-year teachers and principals had different perceptions of the elements necessary for teaching effectiveness. Both groups of teachers brought expectations to the first year of teaching that were not met. Similar support services were offered to both groups, though not at a high level for either group.	Yes, t-tests were conducted.	None reported, but effect sizes could be computed from data provided.
Klug & Salzman (1991)	Mean difference scores calculated to compare pretest and posttest values on teacher competencies and attitudes; one-way ANOVAs.	Group 1 (strong induction program) had greater positive increases compared with Group 2 (weak induction program) on five of the subscales of the questionnaire: rapport with the principal, curriculum issues, teacher status, community support, and community pressures; there was a negative association with attitudes about school facilities. Ratings of teacher competencies slightly decreased for both groups. Novice teachers in both groups had lower post-induction scores on two subscales of the questionnaire: teacher salary and school facilities.	Yes, tests of statistical significance were conducted and reported.	None reported, but effect sizes could be computed from data provided.
New York City Board of Education (1993)	Frequencies, percentages, means (no standard deviations), t-tests, two-way ANOVAs, and chi-squares were reported.	Longer mentoring was associated with improvement in the interns' ability to teach and manage the classroom (based on mentor assessment) though no contrast made with comparison group of new teachers, and on the interns' comfort level with discipline/managing the classroom, writing lesson plans, working with fellow faculty, and confidence in their teaching. Mentored interns seemed to improve more and were more likely to remain in the teaching profession, especially if the mentor was of the same grade and/or subject level. Also, retention was higher for program participants who were assigned mentors at the beginning of the school year rather than mid-year.	Yes, tests of statistical significance were conducted and reported.	Cannot compute from data provided.

Reference	Type of Analysis	Results	Adequate Reporting of Statistical Tests	Accurate Estimates of Effect Sizes
Ratsoy et al. (1987)	Means and t-tests reported.	There were no statistically significant differences in mean competencies of interns and non-interns. At posttest in 1986, the beginning teachers who had been interns in 1985 had higher mean values on 22 of the 26 teaching strategies, compared with their pretests (i.e., at the beginning of the internship year). Teachers who had not been through internship also improved in the second year on 22 of the 26 strategies. (Again, this component of the study did not address whether one group changed more than the other.) Former interns who were first-year teachers in 1986 had significantly higher mean scores on 21 of the 26 strategies than the group that had been first-year teachers in 1985 with no internship participation. Former-intern means were also higher than those for 1986 first-year teachers, though the differences were significant in only five of the 25 strategies.	Yes, tests of statistical significance were conducted and reported.	Cannot compute from data provided.
Schaffer, Stringfield, & Wolfe (1992)	ANOVAs tested pre- to-posttest change within groups.	In their first year, beginning teachers started below the effective percentage for academic statements and improved over the year to achieve on average 80% academic statements. They began above the effective percentage of 12% for organization statement, and while they made significant improvements (dropping from 22% to 15%) they ended the year still above the target value. There was no change in their behavior statements, which at 3.8% remained somewhat above the target value. Teachers who remained for a second year showed a pattern for achievement and organizational statements in which they started the second year at a level better than their end point of the first year and declined somewhat over the year, but remained still improved over their first-year post performance. In behavioral statements, the teachers improved (by making fewer statements) over the second year and achieved the target goal of < 3%, on average.	Reporting was adequate.	None reported, but effect sizes could be estimated from the data provided.
Stallion (1988)	ANOVAs for the observation data; however, no statistics or numerical data are reported, only text summary. The analysis compared two groups at a time rather than a one-way ANOVA across the three groups. Chi-squares on frequencies were used to examine conference report data.	The two trained groups did not differ on any of the measures. The untrained group differed from each of the trained groups in the same ways: untrained group had more "off task" behaviors in their classrooms, and teachers discussed management problems more often with their mentors. No differences between control and instructed groups on the other classroom observation measures, which were ratings by the observer.	No data were provided; significance tests summarized in the text.	Cannot compute from data provided.

Reference	Comparability of Treatment & Control Groups	Possible Confounding Factors	Generalizability of Findings -- Inclusive Sampling	Generalizability of Findings -- Interaction Effects	Additional Notes
Brown & Wambach (1987)	No information was provided on how teachers were selected or identified for the groups nor information that would allow one to examine comparability of groups.	Lack of comparability of treatment and comparison teachers. Contamination, attrition not discussed and cannot be addressed.	Unknown. Method of selecting sample and comparison groups is not described. No information about grade levels taught, teaching experience of mentors, or content of treatment.	Some attempt to look at effect for student teachers vs. first-year teachers.	
Cheng & Brown (1992)	Comparison group was not necessarily comparable to the treatment group in 1991-92 since a larger proportion of teachers in the comparison group had prior experience teaching compared with the treatment group.	The comparison group was composed differently for each year of the evaluation. In the first year, teachers who applied to participate in the induction program but did not meet the criteria for selection were assigned to the comparison group; in the second year, since all interested participants were accepted into the program, a randomly selected group of elementary teachers was assigned to be the comparison group. Also, in year two, the comparison group had a larger proportion of teachers with prior teaching experience than the treatment group.	Demographic data were comparable for the treatment and comparison groups in 1990-91; however, there are differences in 1991-92 based on how the comparison group was composed. Also, it is not clear whether the sample reflects all first-year teachers in Toronto schools (the target population).	No interactions were reported.	Data also collected and reported about mentor teachers.
Charles A. Dana Center (2002)	Demographic makeup of TxBESS teachers similar to the population of non-TxBESS teachers; however, TxBESS serves a greater percentage of teachers in economically disadvantaged areas.	It was not possible to keep track of all the participants through the database, and some regions had different names for the program. Comparable groups for the student achievement analysis could not be built because few districts had both control teachers and teachers who participated in the program at the grades tested.	Generalizable to Texas.	No interactions were reported.	
Gold (1987)	No information provided to evaluate this or to assess if random assignment occurred. Some indication that reassignments had to be made when a teacher was resistant to accepting assistance from the mentor assigned.	Attrition dropped overall during the year of the study because of substantial increase in salary schedule for new teachers, though there is no information provided on differential attrition. Mentors and mentees were in the same schools as the control group teachers.	Report lacks information on the characteristics of teachers, mentors, and schools, so this is difficult to evaluate. The study is weak in that there is only one outcome measure.	No interactions were reported.	

Reference	Comparability of Treatment & Control Groups	Possible Confounding Factors	Generalizability of Findings -- Inclusive Sampling	Generalizability of Findings -- Interaction Effects	Additional Notes
Gunter (1985)	Unclear. On the one hand, matching by demographic features is a strength; on the other, it is not clear how, if at all, first-year teachers who applied for certification before the new law (adding the program as a requirement) might differ from those who did not apply until after the law was put in place.	No information was provided about possible contamination since both types of teachers could be in the same school, nor is there information about the extent to which teachers' first-year experiences differed based on program participation since "nonparticipants" were also observed during their first year of teaching.	Participants and nonparticipants taught at three different districts in Florida; the sample represented a range of grade levels.	No interactions were reported.	
Henry (1988a)	No evidence was provided to assure the comparability of the treatment and control groups. It is unclear if there were systematic differences between teachers in the treatment and control groups since the treatment group was composed of the first 20 new teachers employed in the 10 participating school districts.	Treatment group was composed of the first 20 new teachers hired in 1986-87 in the 10 school districts participating in Project CREDIT, so it is possible there were differences between this group of beginning teachers and others who were hired later and potentially were selected to be in the control group.	Unknown since characteristics were not provided about the sample of beginning teachers who participated in the evaluation study.	No interactions were reported.	
Kilgore & Kozisek (1989)	Though the control group teachers met the same program entry requirements as the experimental group, they chose, for one reason or another, not to participate in the induction program.	Control group teachers had been accepted to the induction program yet chose not to participate; no information provided to assess whether this makes this group of beginning teachers different from the experimental group that did choose to participate.	Not clear what the program entry requirements were (i.e., criteria for inclusion in the research). Since the induction program is a selective one, it is unlikely that the participants are representative of all beginning teachers, though no data were provided to test this assumption.	No interactions were reported.	Qualitative data analyses and results are not discussed in this review.
Klug & Salzman (1991)	Inadequate information to assess the comparability of the two groups, one participating in a strong induction program, the other in a weak induction program.	The authors noted that videotaping might have affected measures of teacher competency if teachers were not comfortable with the technique. Also, qualitative data suggest that teacher competencies increased, contradicting the scoring from videotapes.	No differences (or lack thereof) were reported between first- and second-year teachers. The relatively rural setting for a substantial portion of the sample may be an issue for generalizability.	No interactions were reported.	
New York City Board of Education (1993)	The authors described the treatment and comparison groups in terms of degree attainment, teaching experience, certification, and grade level taught; though they likened the groups on a number of characteristics, it is not clear from the data presented that these two groups are comparable.	Not all of the programs started at the beginning of the school year in 1992; most began in January 1993; the survey was sent in May 1993. Low survey response rates reported: 41% of intern group, 22% of the control group.	Should be able to generalize to other new teachers in NYC.	No interactions were reported.	

Reference	Comparability of Treatment & Control Groups	Possible Confounding Factors	Generalizability of Findings -- Inclusive Sampling	Generalizability of Findings -- Interaction Effects	Additional Notes
Ratsoy et al. (1987)	Seems defensible, though it is not clear how internship participants were selected. Authors mentioned that there were no systematic differences between interns and non-interns in their schooling. Given that first-year teachers had higher GPAs and higher practicum grades than interns, any findings that interns performed more competently than non-interns would be particularly noteworthy.	There is no mention of whether observers were blind to the status of the person they were observing (intern vs. non-intern).	Treatment in this case is an internship that does not involve full-time classroom responsibilities. Thus, findings would not necessarily generalize to full-time teachers in an induction program. It is not clear if people who obtained a job in year 2 of the evaluation were systematically different in any way from those who did not obtain a teaching job that year; hence, the value of the internship for those who did not obtain a teaching position is not known.	No interactions were reported.	Only the "Classroom Observation Study" component of a larger project is reported here.
Schaffer, Stringfield, & Wolfe (1992)	No adequate comparison group of beginning teachers was included.	Since there was no comparison group of beginning teachers, it is not possible to distinguish change in teaching behavior due to the treatment from the natural, developmental change that occurs over time in the first year of teaching. Part of the treatment was to share information about the pre-levels of classroom behavior gathered through the observation instrument. Since there were no other independent measures of teaching practices, nor a control group, it is not possible to evaluate the effect of this practice on teachers' posttest scores.	The findings are limited to beginning teachers who were interested, and qualified to participate, in a two-year master's program. Treatment may have been more intensive and selective than many induction programs can provide; also, participants had slightly higher GPAs than those who were not selected.	No interaction effects for beginning teachers were tested. Contrasts with change in experienced teachers provide some information about the generalizability of the global treatment, teacher training in effective use of class time.	Data on experienced teachers not summarized here.
Stallion (1988)	Good -- random assignment of teachers to groups.	The sample included not only beginning teachers but also returning teachers and teachers changing grade levels or subject areas. Treatment took place mid-year. Timing of observations during the semester following treatment was unclear.	Multiple measures of classroom management, observations taken four times over the semester and averaged, though not clear whether one semester was long enough for effects to take place since the intervention occurred mid-year.	No interactions were reported.	

Appendix B:
**References for Experimental, Quasi-experimental, and Correlational Research and
Reviews of Literature**

- Bainer, D. L., Kramer, G. A., & Smith, R. M. (Eds.). (1999). *Mid-Western Educational Researcher, Volume 12*.
- Blair-Larsen, S. M., & Bercik, J. T. (1990). *A collaborative model for teacher induction*. (ERIC Document Reproduction Service No. ED341642)
- Bracht, G., & Peters, M. (1989). *Minnesota's Teacher Mentorship Program. Formative evaluation report, September 1989. Submitted to the Minnesota State Legislature in fulfillment of M.S. 125.231*. St. Paul, MN: Minnesota State Department of Education.
- Brooks, D. M. (1986). *Richardson New Teacher Induction Program. Final data analysis and report*. Richardson, TX: Richardson Independent School District.
- Brouillet, F. B. (1987). *Report to the legislature on the Beginning Teacher Assistance Program (BTAP). 1985-1986*. Olympia, WA: Washington Office of the State Superintendent of Public Instruction.
- Brown, J. G., & Wambach, C. (1987, February). *Using mentors to increase new teacher retention: The Mentor Teacher Induction Project*. Paper presented at the annual meeting of the American Association of Colleges for Teacher Education, Arlington, VA.
- Cabello, B., et al. (1995). The Comprehensive Teacher Institute: Successes and pitfalls of an innovative teacher preparation program. *Teacher Educator, 31*(1), 43-55.
- California Commission on Teacher Credentialing & California State Department of Education. (1992). *Success for beginning teachers. The California New Teacher Project 1988-92*. Sacramento: Author.
- Charles A. Dana Center. (2002). *Texas Beginning Educator Support System: Evaluation report for year three, 2001-02*. Austin: University of Texas, College of Natural Sciences.
- Cheng, M., & Brown, R. S. (1992). *A two-year evaluation of the Peer Support Pilot Project: 1990-1992*. Toronto, Ontario, Canada: Toronto Board of Education, Research Department.
- Chiang, L. H. (1989, October). *The impact of mentoring on first year teachers*. Paper presented at the annual meeting of the Mid-Western Educational Research Association, Chicago, IL.
- Cousin, S. L. (2001). *An analysis of stress factors and induction practices that influence a novice teacher's intention to stay in the profession*. Unpublished doctoral dissertation, University of Southern Mississippi.
- Dalton, S. S., et al. (1994). *A program evaluation of support-and-assessment of beginning teachers: A responsive-teaching alternative to diagnosis-and-prescription*. (ERIC Document Reproduction Service No. ED377176)

- Davis, B. H., Resta, V. K., Higdon, K. A., & Latiolais, L. L. (2001). Teacher Fellows: A graduate program for beginning teachers. *Action in Teacher Education*, 23(2), 43-49.
- Durbin, D. (1991). *Review of pre-1980 and post-1980 induction programs*. (ERIC Document Reproduction Service No. ED338562)
- Eberhard, J., Reinhardt-Mondragon, P., & Stottlemyer, B. (2000). *Strategies for new teacher retention: Creating a climate of authentic professional development for teachers with three or less years of experience*. Corpus Christi: Texas A&M University, South Texas Research and Development Center.
- Feiman-Nemser, S. (1996). *Teacher mentoring: A critical review* (ERIC Digest No. EDO-SP-95-2). Washington, DC: ERIC Clearinghouse on Teaching and Teacher Education.
- Feiman-Nemser, S., Schwille, S., Carver, C., & Yusko, B. (1999). *A conceptual review of literature on new teacher induction*. Washington, DC: National Partnership for Excellence and Accountability in Teaching.
- Freiberg, M., et al. (1994, February). *Perceptions of beginning teachers in an urban setting: Does mentoring make a difference?* Paper presented at the annual meeting of the Association of Teacher Educators, Atlanta, GA.
- Giebelhaus, C., & Bendixon-Noe, M. (1997). Mentoring: Help or hindrance? Research alive. *Mid-Western Educational Researcher*, 10(4), 20-23.
- Gold, M. (1987). *Retired teachers as consultants to new teachers: A new inservice teacher training model. Final report. Case 09-87*. New York: City University of New York, New York Institute for Research and Development in Occupational Education; Washington, DC: American Association of State Colleges and Universities.
- Gunter, M. N. (1985). *A comparative study of participants and nonparticipants in the Florida Beginning Teacher Program*. Fort Myers, FL: Florida Educational Research and Development Council.
- Harris, M. M., Holdman, L., Clark, R. E., & Harris, T. R. (2001, October). *Rural teachers in Project Launch*. Paper presented at the Research Forum of the National Rural Education Association, Albuquerque, NM.
- Heck, R. H., & Blaine, D. D. (1989, March). *The effects of a colleague-pairing induction program on the perceptions of first-year teachers and their mentors*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.
- Hendrick, L. S., & Childress, L. J. (2002, April). *The RIMS Beginning Teacher Support and Assessment Partnership: A study of eight years of collaboration*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- Henry, M. A. (1988a). *Project CREDIT. Certification Renewal Experiences Designed to Improve Teaching. A cooperative multiple-support program for first-year teachers between Indiana State University and ten public school districts*. Terre Haute: Indiana State University, Department of Secondary Education.

- Huffman, G., & Leak, S. (1986). Beginning teachers' perceptions of mentors. *Journal of Teacher Education, 37*(1), 22-25.
- Hulick, C., & Malone, B. G. (1988). *The Kentucky Beginning Teacher Internship Program: A preliminary evaluation*. Paper presented at the annual conference of the National Council on Inservice Education, New Orleans, LA.
- Huling-Austin, L. (1988, April). *A synthesis of research on teacher induction programs and practices*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- Huling-Austin, L., et al. (1989). *Assisting the beginning teacher*. Reston, VA: Association of Teacher Educators.
- Huling-Austin, L., & Murphy, S. C. (1987, April). *Assessing the impact of teacher induction programs: Implications for program development*. Paper presented at the annual meeting of the American Educational Research Association, Washington, DC.
- Hunter, J. (1988). *Induction of new teachers: An annotated bibliography*. (ERIC Document Reproduction Service No. ED304428)
- Ingersoll, R., & Kralik, J. M. (2003, June). *A review of empirical research on the effects of teaching mentoring programs in elementary and secondary schools*. Unpublished paper written for the Education Commission of the States by the Consortium for Policy Research in Education, University of Pennsylvania.
- Jensen, M. C. (1986, September). Induction programs support new teachers and strengthen their schools. *Oregon School Study Council Bulletin, 30*(1).
- Kilgore, A. M., & Kozisek, J. A. (1989). The effects of a planned induction program on first-year teachers: A research report. In J. Reinhartz (Ed.), *Teacher induction. NEA aspects of learning*. Washington, DC: National Education Association.
- Klug, B. J., & Salzman, S. A. (1991). Formal induction vs. informal mentoring: Comparative effects and outcomes. *Teaching & Teacher Education, 7*(3), 241-251.
- Knauth, S., & Kamin, C. (1994, April). *What happens when it's not training or induction but training and induction*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- Maline, M. B., (1994). *Teacher development outside the United States. A selected annotated bibliography. Recent works collected by the Office of Research, U.S. Department of Education*. Washington, DC: Office of Educational Research and Improvement, Office of Research.
- McDonald, F. J. (1982). *Study of induction programs for beginning teachers. Executive summary*. Princeton, NJ: Educational Testing Service.
- Mitchell, D. E., Scott, L. D., Hendrick, I. G., & Boyns, D. E. (1998). *The California Beginning Teacher Support and Assessment Program. 1998 statewide evaluation study*. Riverside, CA: University of California Riverside School of Education, California Educational Research Cooperative.

- New York City Board of Education, Office of Research, Evaluation, and Assessment. (1993). *Mentor Teacher Internship Program, 1992-93. OERA report*. Brooklyn, NY: Author.
- Odell, S. J., & Ferraro, D. P. (1992). Teacher mentoring and teacher retention. *Journal of Teacher Education*, 43(3), 200-204.
- Olebe, M. G. (2001). Can state policy mandate teacher reflection? *Issues in Teacher Education*, 10(2), 9-21.
- Price, J. L., et al. (1989, November). *A state entry-year teacher program and teacher education*. Paper presented at the annual meeting of the Mid-South Educational Research Association, Little Rock, AR.
- Quisenberry, N. L., et al. (1989). *Teacher induction: An annotated bibliography, 1985-89*. (ERIC Document Reproduction Service No. ED317517)
- Ratsoy, E., et al. (1987). *Evaluation of the Initiation to Teaching Project. Final report*. Edmonton, Alberta, Canada: Alberta Department of Education.
- Reinhartz, J. (Ed.). (1989). *Teacher induction. NEA aspects of learning*. Washington, DC: National Education Association.
- Rhoton, J., & Bowers, P. (Eds.). (2003). *Science teacher retention: Mentoring and renewal. Issues in science education*. Arlington: National Science Teachers Association.
- 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(3), 181-192.
- Scott, N. H. (1997). *PFD's for beginning teachers: A report on the teacher induction program in New Brunswick*. New Brunswick, Canada: New Brunswick Department of Education.
- Scott, N. H. (1997). *Beginning teacher induction program in New Brunswick, 1996-97 report*. Fredericton, New Brunswick, Canada: New Brunswick Department of Education.
- Scott, N. H., & Compton, E. (1996). *Report on the 1995-96 beginning teacher induction program in New Brunswick*. New Brunswick, Canada: New Brunswick Department of Education.
- Selke, M. J. G. (1993). *Teacher induction variables: Impact upon second year teacher retention*. Unpublished doctoral dissertation, Marquette University.
- Serpell, Z., & Bozeman, L. A. (1999). *Beginning teacher induction: A report on beginning teacher effectiveness and retention*. Washington, DC: National Partnership for Excellence and Accountability in Teaching.
- Stallion, B. K. (1988, February). *Classroom management intervention: The effects of training and mentoring on the inductee teacher's behavior*. Paper presented at the annual meeting of the Association of Teacher Educators, San Diego, CA.
- Stewart, D. K. (1992). Mentoring in beginning teacher induction: Studies in the ERIC data base. *Journal of Teacher Education*, 43(3), 222-226.

- Thomas, P., & Newton, S. (2001, April). *Evaluating the effectiveness of beginning teacher programs in a large urban school district*. Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.
- Tinsley, S. H. (1986). *Rated severity of problems and helpfulness of the induction process for student teachers and first-year teachers*. Unpublished doctoral dissertation, University of Arkansas.
- U.S. Department of Education. (2000). *Taking the next bold step to America's future: A national conference on teacher quality* (Washington, DC, January 9-10, 2000). Washington, DC: Author.
- Walker, L. (1992, November). *Mentoring: A review of the literature*. Paper presented at the annual meeting of the Mid-South Educational Research Association, Knoxville, TN.
- Washington Office of the State Superintendent of Public Instruction. (1988). *Report to the legislature on the Teacher Assistance Program. (Chapter 507, Laws of 1987)*. Olympia, WA: Author.
- Watson, A. J., Hatton, N. G., Squires, D. S., & Soliman, I. K. (1991). School staffing and the quality of education: Teacher adjustment and satisfaction. *Teaching & Teacher Education, 7*(1), 63-77.
- Weisbender, L. (1989). *Preventing teacher dropout: Volume 1—Mentors helping new LAUSD teachers, 1984-1988; Volume 2—Retention rates for LAUSD mentors, mentees, and other teachers, 1984 to 1988* (No. 531). Los Angeles: Los Angeles Unified School District, Program Evaluation and Assessment Branch.
- Weiss, E. M., & Weiss, S. G. (1999). *Beginning teacher induction* (ERIC Digest No. EDO-SP-1999-3). Washington, DC: ERIC Clearinghouse on Teaching and Teacher Education.
- Whitebook, M., & Sakai, L. (1995). *The potential of mentoring: An assessment of the California Early Childhood Mentor Teacher Program*. Washington, DC: National Center for the Early Childhood Work Force.
- Yosha, P. (1991, April). *The benefits of an induction program: What do mentors and novices say?* Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.
- Zetler, A. C., & Spuhler, L. (1995). *Montana Beginning Teacher Support Program. Research report for year three 1994-95*. Helena: Montana State Board of Education.