RESEARCH GENRES IN TEACHER EDUCATION[1]

Mary M. Kennedy[2]

Almost since its inception, teacher education has suffered from doubts about its value to teachers. Outside observers have asked, usually skeptically, whether teacher education makes a difference, and teacher educators themselves have wondered what they have been able to accomplish and how they could accomplish more. Presumably, research could help both teacher educators and teacher education policymakers to understand better whether and how teacher education makes a difference. But the question has been difficult to answer because the enterprise itself is extremely complex. It is a huge enterprise, producing 100,000 new teachers each year; it occurs in a wide variety of institutions of higher education; and it occurs in other kinds of institutions as well. Even within higher education, there are people enrolled in teacher education who won't become teachers and people not enrolled who will become teachers. Finally, the boundaries between teacher education and not-teacher-education aren't clear. Some of us count experiences in schools as part of teacher education; some count courses in the liberal arts as teacher education; and some count only those courses that occur within education departments as teacher education.

Not only is the enterprise itself difficult to get a handle on, but the outcomes of teacher education are similarly diffuse. As a field, we suffer from enduring disagreements about what counts as a valid outcome and about how to measure those outcomes that do count. Some people want evidence of teacher thinking, others of teacher skills. Some think you assess thinking through paper-and pencil tests, others that you need to see it in the context of practice. And so forth.

The complexity and size of the enterprise, coupled with the ambiguities about what counts as an outcome, make it difficult for researchers to pose manageable research questions and to design studies that can improve our understanding of teacher education and what it does. In order to make their task more manageable, researchers limit their attention to problems that they can easily define. Here are three examples of how researchers delimit their scope of inquiry.

1. *Many researchers concentrate on the student teaching component of teacher education* (Goodman, 1986; Hodges, 1982; Silvernail and Costello, 1983; Tabachnik and Zeichner, 1984). The student teaching component is a more definable and therefore manageable piece of the teacher education puzzle. It occurs in a definable time slot, in a definable place, and is relatively separate from the rest of teacher education. The typical research design for studies of student teaching is a single-group longitudinal design; that is, researchers contrast before and after data on student teachers' beliefs or knowledge or skills. One advantage of these studies is that, since student

---

[1]An earlier version of this paper was presented at the annual meeting of the Association of Teacher Educators, February 1991.

[2]Mary M. Kennedy, professor of teacher education, is director of the National Center for Research on Teacher Learning.

teaching experiences are the least well-controlled aspect of teacher education, it is possible to capitalize on the variation in student teaching experiences to learn more about what features make the most difference to different kinds of outcomes (eg., McIntyre and Killian, 1986; 1987). But this variation is also a disadvantage, precisely because student teaching is the least controlled aspect of teacher education. Moreover, while these studies are valuable, and have increased our understanding of the student teaching component of teacher education, they leave untouched the centerpiece of the enterprise—the large, diffuse, complicated web of courses and other events that we call preservice teacher education.

2. *Other researchers study inservice programs rather than preservice programs* (Carpenter, Fennema, Peterson, Chang, and Loef, 1989; Coladarci and Gage, 1984; Good, Grouws, and Ebmeier, 1983; Griffin and Barnes, 1986). Inservice teacher education programs are more manageable from a research point of view than preservice programs are. They have a clearly defined starting and stopping points and clearly defined groups of program participants. Often, they also have more clearly defined goals: Many of them aren't preparing teachers to do *everything*, but instead are focusing, for instance, on teaching secondary science, or on teaching elementary reading, or on increasing time on task in elementary classrooms. So researchers have a more manageable task when they study inservice teacher education. Yet, although we have learned a lot about inservice teacher education from these studies, such studies do not shed light on preservice teacher education, which continues to be the dominant part of the enterprise.

3. *Still other researchers limit their inquiry to description of parts of the system rather than exploring relationships among parts* (eg., American Association of Colleges of Teacher Education, 1987;1989; Howey and Zimpher, 1989; 1990; Kluender, 1984). Instead of looking at how teacher education programs influence teacher candidates, these researchers look at what teacher education programs are like, what teacher education faculty are like, what their goals are, what their curriculum requirements are, or how student teaching is integrated with course work. These studies help us define this big, complex enterprise we call preservice teacher education, but they do not help us better understand it, for they do not tell us whether any of these dimensions make a difference.

All of these approaches to research in teacher education have been profitable. We know a lot more now than we did even 10 years ago about what happens during student teaching, about how inservice programs work, and about what preservice teacher education programs are like and who teaches in them.

But none of these bodies of knowledge helps us better understand whether or how the central part of teacher education makes a difference. This is not to say that *no* research has been done on the impact of preservice teacher education. Indeed, a variety of approaches have been devised over time to try to get a better handle on preservice teacher education. But because of the complexity of

the enterprise, and because of the ambiguity about its intended outcomes, no study can accommodate all aspects of teacher education and all outcomes. Every researcher necessarily limits his or her attention. Every researcher makes difficult decisions about which aspects of teacher education will be studied, about which outcomes will be examined, and about how the study will be designed to determine the relationship between teacher education and its outcomes.

Though there is, in principle at least, an infinite number of ways these decisions could be made, the available research on preservice teacher education tends to fall into five distinct categories, or *genres*. Each genre represents a particular way of thinking about whether or how teacher education makes a difference. I call them genres because each represents a coherent and internally consistent way of thinking about whether teacher education makes a difference, because each has been used on numerous occasions by numerous researchers, and because researchers within each genre tend to build on other work within their genre more than on work in other genres. Several of them represent communities of scholars who share a set of norms and values, and who share a particular view of, and interest in, teacher education.

Each genre represents a particular way of thinking about what teacher education *does* do and what it *can* do to some important outcome. Each holds a different kind of promise for helping us understand teacher education. But because each has had to delimit its inquiry in important ways, each also is limited in what it can tell us. The differences among these genres is most apparent in the delimiting decisions they make. They differ, for instance, in the *aspects of teacher education* that they choose to examine. Some focus on completed programs, some on particular components within preservice programs, and others on the volume, or number of courses taken, in teacher education. They also differ in the kind of *what counts as an outcome*. Some examine skills; some examine teacher knowledge or beliefs; and still others look for evidence of gains in pupil achievement. And they differ in the *kind of arguments* they make about whether or how teacher education makes a difference.

My aim in this paper is to examine these genres with an eye toward the kind of knowledge each gives us. My hope is that, through such a review, teacher educators and researchers can learn more about how research can contribute to our knowledge and understanding of teacher education, and perhaps can find ways of improving on these genres in the future. As I review each genre, I ask three questions of it. First, what aspects of teacher education does it look at, and are those aspects relevant to the needs of teacher educators who want to use research to improve their programs? Second, what outcomes does it look at, and are these outcomes sufficient? And third, what is the argument about the relationship between these two, and is the argument credible?

Open Searches for Contributions to Student Learning

One way to think about the role of teacher education is to assume that, if teacher education matters, it should make a difference in the achievement of students whose teachers have had different amounts or kinds of teacher education. Researchers working from this assumption are not testing any particular theory about teacher education, or about anything else that might influence student achievement. Instead, they are engaged in a relatively open-ended search for contributions to student learning, and one of the possible contributions is teacher education. Some of the factors that influence pupil learning are within the students themselves—their academic ability, for instance, the language they speak, and their motivation to succeed in school. Other factors are found in the students' families—in the education levels of their parents, for instance, in their family income, or in the number of siblings they have. Still other contributors to student achievement are found in the schools—in the textbooks they use, in the size of their libraries, and in their school policies. And still other contributors to student achievement reside in the teachers themselves: in their verbal fluency, in their education levels, and in the number of years experience they have had as teachers.

Researchers who practice within this genre are interested in the question of what contributes most to student achievement. And they are especially interested in those factors that schools can control. For instance, if they found that library size didn't make a difference, but that class size did, they would advise local school boards to spend their money reducing class size rather than building up their libraries. That is the type of outcome these researchers seek. Their audiences are school district policymakers, not teacher educators. Yet because they examine all the factors that might be relevant to student achievement, they wind up doing research on whether teacher education makes a difference as well.

Numerous researchers have examined contributions to student achievement in the last several decades. Many of these studies were stimulated by, and are based on, the Equality of Educational Opportunity Study (Coleman et al., 1966), and many actually used the EEOS data. One of the earliest and best of these studies that included teacher education was conducted by Eric Hanushek (1971, 1972). He began by asking whether teachers in general differed in their ability to increase student achievement, after taking into account the child's initial achievement and various aspects of the child's background. Hanushek found that teachers did make a difference; that is, the teacher a child happened to have could significantly influence the child's achievement for the school year. Seeing that this was the case, Hanushek then tried to see which particular teacher characteristics seemed to account for these differences. Among the variables Hanushek examined were college major, number of hours of graduate course work teachers had taken, and length of time since the teachers' most recent educational experience. Hanushek found that neither college major nor the number of graduate credits teachers had taken were significantly related to student achievement.

Variables that were related, in contrast, included the teachers' general verbal ability and the recency of their last educational experience. These two variables do not necessarily reflect teacher education courses per se, although verbal ability may reflect the effect of college education in general. The recency of the teachers' educational experiences may reflect either the nature of the experiences or the teachers' interest in continued professional learning.

Another important study that focused on teacher education was done by Murnane and Phillips (1981). Like Hanushek, they began by testing to see whether teachers made a difference to student achievement in general, and found that they did. They then tried to see what teacher characteristics seemed to account for these differences. But instead of generating a single equation which included all possible contributions to student achievement, they developed two separate equations, one of which included measures of teacher *behaviors* and the second of which included measures of teacher *characteristics*. For their first equation, they predicted student achievement using a number of specific teacher behaviors (e.g., circulating around the room to correct seatwork, using demonstrations, making students repeat poor work, etc.). For their second, they tried to predict student achievement on the basis of teacher characteristics such as years of experience, possession of a master's degree, and prestige of college attended. Their data indicated that teacher behaviors were better predictors of student achievement than were teacher characteristics. Moreover, of those characteristics Murnane and Phillips examined, neither of their education-related variables—possession of a master's degree and prestige of college attended—appeared to be relevant to student achievement.

Begle and Geeslin's (1972) study provides an example that looks more closely at the pattern of courses that teachers took. They focused specifically on mathematics teachers and included some 20 different teacher characteristics in their study, including whether the teachers majored or minored in mathematics and the number of course credits teachers took in mathematics. Even with their attention to the undergraduate curriculum, and to the specific subject being taught, they still found little relationship between teacher's course taking and student gains in achievement in mathematics.

In a recent review of literature in this genre, Hanushek (1989) summarized 113 studies that included some aspect of teachers' education. Only 13 of these education-related variables were statistically significant; of these, 8 indicated that the teachers' education was positively related to student achievement and 5 indicated that it was negatively related to student achievement. Unfortunately, Hanushek's summary does not indicate the particular aspects of teachers' education that were measured in these studies. Some studies may have measured whether or not the teacher majored in an academic subject, others may have measured the number of credits taken beyond the bachelor's degree, while still others measured the recency of the education experience. We don't

know which of these were measured or how often any of them was measured in this collection of studies.

Now let me address my three questions about this genre of research.

Aspects of Teacher Education Examined

The aspects of teacher education that these researchers examine are often called policy-parameters: broad parameters of teacher education that can be manipulated by policymakers. These researchers do not ask about the details of any particular teacher education program, but instead ask, for instance, whether teachers majored in education or in some other subject, whether they held bachelor's or master's degrees, how recently they received their educations, and so forth.

Usually, quantifiable measures of these aspects of teacher education are justified on the ground that these measures represent the dimensions of teacher education about which policies are formulated; that is, most state policies require that K-12 teachers must hold a teaching credential and that they must have participated in an accredited teacher education program. And most school districts provide additional salaries to teachers who hold master's degrees. But although these aspects of teacher education are, on their surface, important to policy, they may not be of much real use to those who want to improve teacher education, for two reasons.

First, virtually every teacher in these studies already holds a bachelor's degree and is already certified to teach. The number of teachers lacking a bachelor's degree was only 7 percent in 1966, and has since fallen to less than 1 percent (National Center for Education Statistics, 1989). Presumably, then, all of these teachers have attained the minimum educational background required for teaching. The variations among these teachers that are measured, therefore, are not variations in the most fundamental aspects of teacher education, but instead are peripheral variations. Statisticians refer to this as a problem of *restricted range*. If a group is overly homogeneous, it will be difficult to show a relationship between one variable and another. A wider range is needed in order to see such relationships.

A second reason these studies might lack utility has to do with variations in educational backgrounds that are *not* measured. Since the United States does not have a centralized curriculum, and since many states give teacher educators considerable leeway in their program designs, teacher education programs can look remarkably different from one institution to the next. Teacher education looks different at Doane College than it looks at Swarthmore, and it looks different at Swarthmore than it does at Illinois State University. A recent report from the Council of Chief State School Officers (1988) indicated that the number of credits of professional education required for elementary teacher candidates ranged from 18 to 90 across the states.

These differences reflect different theories and different assumptions about what teachers need to know and about how teachers learn. It is reasonable to suppose that such differences are relevant to the outcomes of teacher education, but they are not differences that can be easily measured. By failing to measure the substantive differences among programs, researchers in this genre may miss the very aspect of teacher education that is most likely to make a difference. Moreover, because there is so much variation in the content and character of teacher education programs, any measure of the *amount* of teacher education will be unreliable;[3] that is, it will not measure a unified or clearly defined phenomenon. Some teachers may have received extensive education in a mediocre program while others received modest education in a very good program. It should not be surprising, therefore, that these measures generally do not correlate highly with measures of student achievement gains.

Outcomes

Researchers practicing within this genre take gains in student achievement as their primary outcome of interest. They have found clear evidence that some teachers promote greater gains in student achievement than others do, and they want to know which teacher characteristics account for these differences.

Some people have argued that, because student achievement can be influenced by many things other than teaching, it is not ethical to use it as a criterion to assess either teachers or teacher education (e.g., Medley, 1982); that is, a teacher may appear to be more or less effective depending on which students the teacher happens to be teaching. But the researchers in this genre are not assessing the ability of any *individual* teacher; instead they are asking whether teachers with certain kinds of college degrees tend to have more or less influence on student achievement. Moreover, when they ask this question, they rely on a statistical technique that is designed to take into account many of the other factors that influence student achievement.

Nevertheless, their reliance on student achievement is a limitation, for there are many things teachers try to accomplish with their students besides raising test scores. These tests measure some, but not all, of the important goals of education. And a good teacher education program will try to help teachers learn to teach their pupils many things other than the basic skills that are measured on standardized achievement tests. Thus, a more appropriate question to raise about this genre is whether pupils' standardized test scores are the most appropriate outcome to use for judging the impact of teacher education. In fact, whether they even measure the most important outcomes of

[3]I am indebted to Willis Hawley for suggesting this point to me.

schooling is a highly debatable issue. Gains in student achievement, then, constitute an overly narrow outcome for estimating the contributions of teacher education to teaching.

Credibility of the Argument

The logic of these studies goes something like this: If teachers who have taken more credits in teacher education foster greater gains in student achievement than teachers with less teacher education (after taking into account differences in entering achievement, family background, and so forth) then teacher education has made a difference. If such differences cannot be observed, we may have reason to doubt the wisdom of policies that require teachers to take a certain number of credits or that pay teachers more if they have master's degrees, for instance. It is a relatively simple argument, but it depends on quite a complex statistical approach called *multiple regression*. Multiple regression is designed not to estimate the effects of any one variable by itself, but instead to weigh individual contributions *relative to* the contribution of other factors that might influence student learning.

What this statistical technique does is create a mathematical model of the set of influences and estimates the relative importance of each. The success of the study depends on how accurately the researcher's *model* of the phenomenon matches the *real phenomenon*. Suppose, for instance, that the researcher develops a model like this:

> Achievement gain = earlier achievement + family support + amount of the teacher's education + recency of the teacher's education.

But that the real phenomenon works like this:

> Achievement gain = earlier achievement + family support + teacher's desire to improve + quality of teacher's education + school climate.

The real phenomenon includes some variables that *differ* from those in the researcher's model—the quality of the teacher's education rather than the quantity of teacher's education, for instance, and the teacher's desire to improve rather than the recency of the teacher's education—and it includes one variable that is *missing* from the researcher's model: school climate. When such differences exists, the study is said to be based on a *misspecified model*, and misspecification can result in two important problems.

The first problem occurs when a variable that has not been measured is correlated with one that has been measured. For example, Hanushek (1971) found that the recency of the teachers' last educational experience was associated with gains in student achievement. One interpretation of this finding is that, in order to continue teaching well, teachers need to continue their education. Either they forget what they learned originally, and therefore need to return to school to relearn it, or teacher educators continue to develop new ideas about teaching and teachers need to return to school to learn the latest ideas. But another interpretation is that the teachers who have taken courses more

recently are the teachers who are more interested in improving their practice anyway. If this is true, teachers who have taken courses recently might be more likely to do better even if they hadn't taken these courses. Because the researcher's model is misspecified, it may lead to the erroneous conclusion that the courses themselves, rather than the teacher's disposition to improve her practice, were responsible for these gains in student achievement.

The second problem deriving from misspecified models occurs when an important contributor is not measured and is also not correlated with one of the variables that is measured. For instance, there is ample evidence now that school climate is an important contributor to student achievement (Good and Brophy, 1986). And it may also be an important contributor to the teachers' ability to teach well. Since the models used by researchers in this genre do not include school climate, variations in school climate add "noise" to the equation, and make all other variables seem less correlated with the outcome than they might really be. It is possible that teachers' educational backgrounds make a difference *within a given school climate*, but these effects are not apparent when a wide range of school climates are involved in the study. Or, it is possible that teacher education only makes a difference within reasonably positive school climates and that it cannot help teachers teach better when they are working in especially difficult schools. When school climate is not included in the researcher's model, its influence on student achievement cannot be known, of course. But more importantly, the researcher cannot know the ways in which school climate may mitigate the influence of other variables, such as teacher education. In this case, misspecification does not yield a false positive relationship but instead yields false negative relationships.

If we are seeking research that can help us reform and restructure teacher education, then, this research genre is limited in all three of the areas we are examining. The aspects of teacher education that it measures miss the essential features of the core undergraduate program, and its outcomes represent only a narrow slice of the outcomes we may want to see. Finally, to the extent that arguments about the relationship between teacher education are based on misspecified models, they lack credibility. And there are at least two plausible ways in which many models are misspecified: They do not take into account the important influence of the teachers' own disposition toward self-improvement and they do not take into account variables such as the character and quality of the teachers' undergraduate program and the climate of the school in which the teacher and students work. Failure to include the first variable can lead to the erroneous conclusion that other associated variables are important (a false positive) and failure to include the latter variables can lead to the erroneous conclusion that teacher education does not make a difference (a false negative).

Comparing the Haves and the Have-Nots

The second way of thinking about the role of teacher education is to compare teachers who have had teacher education with those who have not. Researchers working within this genre contrast practicing teachers who are fully certified with those who are teaching with emergency or provisional credentials. Like the first genre, this one focuses on teachers who have completed their education and are already teaching and then looks back to see what their education was. Usually, researchers focus on teachers within a particular school district or geographic region, find all the teachers who are teaching with provisional or emergency credentials, and then compare them with a sample of teachers in the same region or district who have completed the full complement of required teacher education courses and have become certified. Once two groups of teachers have been identified, the researchers observe the classroom practices of both groups to see whether differences exist in their practices. These studies are difficult to do, in part because they can be done only during periods when school districts are experiencing serious personnel shortages so that they need to hire a great deal of provisionally certified teachers.

The most important way in which these studies differ from one another is in what they observe about teachers and how they do their observations. Studies conducted in the 1960s and 70s tended to use "high-inference" observation instruments—instruments that ask the observer to make judgments about whether, for instance, the teacher is maintaining order or is friendly or aloof with pupils. More recent observation systems tend to rely on "low-inference" devices, in which the degree of observer judgment is severely curtailed. In these studies, observers simply check whether they observed a particular behavior or not but make no judgments as to whether that behavior indicates orderliness or friendliness or any other general teaching trait.

One of the earliest and best examples of this genre is Lupone's (1961) comparison of elementary teachers in New York. This study used a high-inference observation system, and Lupone controlled for differences in observer judgment by using multiple observers in each classroom. In addition, Lupone took into account the number of years of teaching experience his teachers had by grouping them according to whether they had one, two, or three years of experience. Lupone found that fully certified teachers surpassed provisionally certified teachers, across all levels of experience, on four of his seven observation scales. The scales on which teacher education made a difference included preparation and management, subject matter, pupil-teacher relations, and evaluation. On a fifth scale, describing instructional material and methods, Lupone found no difference between first-year teachers, but did find differences between teachers in the other experience categories. The two scales on which no differences were found were parent-teacher relations and human relations, both skills that are demonstrated outside the classroom.

Dewalt's and Ball's (1987) recent study illustrates a low-inference observation system. These researchers compared teachers in Virginia on Virginia's mandated competency assessment. One group of teachers had taken no credits in teacher education, the other had had at least 12 credit hours in teacher education but had not done student teaching. So the comparison really asks whether course work in teacher education makes a difference. The teaching behaviors that were documented through the observation system had been demonstrated in research literature to be effective teaching strategies. When the researchers observed the teachers, they specifically asked their teachers to demonstrate these competencies. Thus, their observations do not reflect what teachers might normally do in their classrooms but instead reflect the teachers' ability to do these specific things on demand.

The two groups were found to differ on several variables, but the comparisons differed in terms of which group was favored. Behaviors that were more often demonstrated by teachers who had taken teacher education courses were those having to do with creating a nonpunitive classroom climate and accommodating individual differences. Those that favored teachers who had taken no courses in teaching had to do with holding students accountable for their work and asking a wide range of questions about the material. These researchers also found, incidentally, a wider range of practices among the nonprepared teachers than among the prepared teachers.

Several recent studies have extended this genre to include comparisons of teachers who participated in alternative routes. For instance, Brown, Edington, Spencer, and Tinafero (1989) compared emergency-permit teachers with both fully certified teachers and interns who were participating in an alternative route program. They pooled their data across grade levels and found no differences among the three groups on four of the five scales they used. Emergency-permit teachers were significantly higher on the fifth scale, called "growth and responsiveness," a scale which could reflect a higher degree of on-the-job learning among these teachers who have received no advance preparation for their work.

Now consider my three questions.

Aspects of Teacher Education Examined

Whereas the open-search researchers are likely to tally up courses or degrees *beyond* the bachelor's degree or to determine whether teachers majored in education or not, researchers in the comparison genre usually define teacher education as the configuration of courses that is required for initial certification. The aspect of teacher education that is of interest to them is the completed program compared to an incompleted program. Since, presumably, these programs are designed to make a difference to teaching practice, the merits of the completed program are of interest. And comparisons among different types of programs—for instance, alternative routes versus traditional

programs—are also of interest, particularly in the current policy climate, where numerous efforts are under way to devise alternatives to the traditional preservice program.

Moreover, many of these researchers also look at the number of *undergraduate* education courses taken by teachers in the noncertified group. An important finding from this research is that very few provisionally certified or emergency-certified teachers have had absolutely no exposure to teacher education. Instead, they have taken a few courses, but not enough to become certified. So the comparisons are actually between teachers who have taken everything that is required to become fully certified and teachers who have taken some portion of the requirements. Again, this difference is relevant, since we design our programs to be sufficient as whole programs.

Still, these researchers still treat the undergraduate program as a black box. While they tally up the number of courses taken, they do not document which courses were taken or from which institution. Nor do they document anything about the nature of those courses. A nice exception to this general rule is Arch's (1989) comparison of teachers prepared through a traditional undergraduate program versus a master's in teaching program. Because both programs were offered by the same institution, Arch was able to examine her teachers' capabilities in light of the specific characteristics of the two programs. Without such an in-depth examination, findings from comparison studies cannot contribute much to reform efforts in teacher education.

Outcomes

Although a few comparison studies use tests of knowledge, such as the NTE or a state-specific required test (e.g., Cornett, 1984), most depend on observations of teachers for their outcomes. But even within the observation studies, there is still a great deal of room for variability in what counts as evidence. These studies have relied on a variety of different observation systems and a variety of different outcomes, depending on what is fashionable at the time and on what observation instruments are available at the time. One could argue, of course, that even though the criteria used in comparison studies may change over time, each criterion is likely to reflect views of good teaching that would also appear in teacher education programs at the time the studies were done. Thus, despite the variability in observation systems, it is reasonable to expect certified teachers to performance better than noncertified teachers in most of these studies.

Credibility of the Argument

I should point out here that there are two very different incentives that guide this research genre. While some researchers are looking for evidence that teacher education has enhanced teaching, others are looking for evidence that it has *hindered* teaching. This second group of researchers view time spent in teacher education as time taken away from courses in arts and

sciences. So comparison studies actually represent a two-sided argument. On one side, if we find greater skill among provisionally certified teachers, we might argue that teacher education hinders teaching and that teachers are better off taking more liberal arts courses than they are taking teacher education courses. On the other side, if we find greater skill among certified teachers, we might argue that teacher education contributes to teaching.

But there are serious limitations to both sides of this argument, for all of these studies examine teaching practice *after* teachers begin teaching. Like the open-search studies, they cannot determine what these teachers were like when they were still in college, making decisions to enter or not to enter a teacher education program or making decisions to take a few courses but not to complete the program. If people with different patterns of capabilities choose these two curricular paths in the first place, the differences we observe when they are teaching could reflect nothing more than the differences that were already there years earlier. Thus a major problem with these studies is that neither group of researchers—those who look for benefits from teacher education or those who look for drawbacks of teacher education—can be sure that the observed differences reflect the courses teachers took.

In fact, not even a finding of *no difference* avoids this dilemma, for it is possible that different kinds of people enroll in different programs and that the programs washed out the initial differences. An interesting study by Skipper and Quantz (1987) illustrates this point. They followed a group of arts and sciences students and a group of teacher education students from their freshman year through their senior year. They found that substantial differences existed between the two groups as freshmen, but that these differences had disappeared by the time the groups were seniors. No difference at the end of a program, then, means no evidence that teacher education has hindered teaching, no evidence that teacher education has contributed to teaching, and no evidence that different kinds of people enroll in different programs to start with.

Beery's (1960) study also illustrates the problem of interpretation. Beery found that certified teachers differed more often from teachers who had taken *some* courses in teacher education than they did from teachers who had taken *no* teacher education courses. Why would such a pattern exist? One strong hypothesis is that the teachers who formed these different groups differed in important ways that may have led them to take the particular configuration of courses they did, so that the differences Beery observed had more to do with what kinds of people chose these curricular paths than with the courses they actually took.

Overall, then, comparison studies focus on a more relevant aspect of teacher education—completed programs—than open-search studies, and their outcomes are more relevant as well. But comparison studies suffer a logic problem that is very similar to that of open searches,

in that neither research genre can separate out the courses or whole programs teachers took from their reasons for taking those courses or programs.

Ask the Teacher

The third way to think about the role of teacher education is to assume that teachers themselves might be the best source of evidence. Teacher educators often try to determine whether particular aspects of teacher education made a difference by surveying their own graduates and asking them if their program made a difference. This strategy is popular in part because it is relatively inexpensive and simple to do, and in part because the National Council for the Accreditation of Teacher Education accreditation requirements have continually stressed the need for such program evaluations. Adams and Craig (1983) surveyed teacher education programs in 1980 and found that 74 percent claimed to be conducting some sort of follow-up of their graduates.

Ask-the-teacher studies generally use two strategies to estimate the contributions of teacher education. One is to ask teachers to assess their own knowledge and skills—that is, to assess their own ability to teach. The other is to ask them to assess the contributions of their preservice program, or the contribution of particular courses within that program, to their teaching.

In 1975, Pigge (1978) surveyed graduates of Bowling Green University and gave them a list of 26 competencies on which the respondents were to rate themselves. On this scale, a rating of 1 meant not proficient and a rating of 5 meant extensive proficiency. The lowest mean score for all 26 competencies was 2.32, a score falling between "limited" and "adequate" proficiency. Teachers felt they were at least adequate on 14 of the 26 proficiencies. Pigge also asked teachers how important these various competencies were to their work and where they learned these competencies. Generally speaking, teachers thought that those competencies *most* necessary to their work were learned on the job, whereas those considered *least* necessary were acquired in their teacher education programs.

Marvin Henry (1986) surveyed the 1983 and 1984 graduates from Indiana State University, asking them to rate themselves on a 3-point scale: "strong," "adequate," and "needs improvement." These were beginning teachers, who presumably should not have been embarrassed to say that they needed improvement on some aspects of teaching. Yet, of the 45 dimensions Henry asked about, only 5 were areas in which 15 percent or more teachers felt they needed improvement. On 9 of these 45 items, no one claimed to need improvement. Henry also asked his beginning teachers whether any of five forms of assistance would be helpful in their beginning years of teaching. The most-often selected options were other teachers or a newsletter. The option least-often selected was "university supervision similar to that received during student teaching."

A prominent part of many institutionally based studies is a list of the specific courses or program components required by the program and a request that the teacher rate the quality or relative value of each part of the program (e.g., Drummond, 1976; Reed, 1975; Schmelter, (n.d.)2; Warren, Dilts, Thompson, and Blaustein, 1982). If student teaching is included in the list, it is invariably the highest rated part of preservice teacher education, usually followed by one or more methods courses. If subject matter preparation is included in the list, it receives a higher rating than professional courses do. If something called *Foundations* is included, or a course with a title like *School and Society* it receives the lowest rating.

In an interesting study by Clark, Smith, Newby, and Cook (1985), teachers were observed in their classrooms and then asked where they got the ideas for what they did. The most frequently cited source for a teaching idea was that the teacher generated it him- or herself. Second most prominent was the cooperating teacher with whom the teacher had undergone student teaching. Teacher education faculty were given credit for only 17 percent of the practices teachers were asked about.

Though most studies are conducted by teacher education institutions and include only graduates of those institutions, a few studies are conducted of teachers in general. For instance, the National Education Association surveyed its members and asked them to evaluate the contributions of 14 different sources of knowledge about teaching, one of which was preservice teacher education (Smylie, 1989). The preservice teacher education program was ranked 13 of 14. The highest rated sources of knowledge were direct experience, consultation with other teachers, and independent study and observations of other teachers, all of which are entirely in the control of teacher him- or herself. The only item rated less positively that undergraduate teacher education was school-district provided inservice programs.

Now consider my three questions.

Aspect of Teacher Education Examined

Because they provide information about the particular components within the program rather than treating the program as a black box with no details illuminated, these studies can be far more informative to the teacher educators than either of the first two research genres are. But their benefit is highly localized: Most of these studies examine teacher education as it exists in one particular institution. They ask their graduates to assess Education 312, the math methods course, the student teaching component, the placement service, and so on. Though many of these components are indeed similar across institutions, the phrasing of survey questions rarely enables faculty from other institutions to understand the significance of the findings, and consequently it is close to impossible to compare findings from one study to the next or to aggregate the findings and identify patterns

regarding different features of preservice teacher education. So even though the aspects of teacher education that they examine are locally relevant, they often do not help the field in general.

Outcomes

Almost universally, ask-the-teacher studies use teachers' judgments of their own knowledge or skill. Most of them provide the teacher with a list of knowledge or skill areas or a list of program courses and ask the teachers to rate themselves or their alma mater on a 5-point scale. A rating of 5 means, "I am highly capable in this area," or "the program was very effective in this area," and a rating of 1 means, "I am extremely incapable in this area" or "the program was extremely ineffective in this area."

Veenman (1984) recently reviewed follow-up survey literature and included studies in other countries as well as those done in the United States. He found that classroom discipline was most often mentioned as a major problem and was nominated in the bulk of the studies he reviewed. The second most often cited problem was motivating students, mentioned in 48 studies; and third most often mentioned was dealing with individual differences, cited in 43 of the studies. From findings such as these, we can distinguish those areas in which teachers feel relatively more capable from those in which they feel relatively less capable. And the areas in which they feel less capable tend to be those having to do with their moment-to-moment interactions with students. Thus, though it is not possible to draw many inferences about teacher education programs, it is possible to learn what teachers think they can do well and what they think they cannot do well.

But the reliance on teacher judgment as an outcome is a substantial limitation in these studies, for several reasons. First, we don't know what *criteria* teachers use when they make these assessments. When a teacher rates herself as adequate or better than adequate, for instance, on what basis does she make this judgment? Are the teachers' criteria the same as an independent observer's criteria might be? Similarly, when a teacher claims a program has contributed to her knowledge or skill, or has *not* contributed to her knowledge or skill, we don't know how accurate these judgments are. It is highly likely that teachers do not recall what they knew or were able to do five years earlier. Strang, Badt, and Kauffman (1987) provide some evidence to support such a process. In their study, they measured teachers' skills both before and after a program treatment, but they also asked teachers afterward to estimate the degree to which they had changed. The researchers' independent assessment of teacher change showed their proficiency moving from 52 percent to 87 percent. However, the teachers' assessments of their change indicated movement from 81 percent to 85 percent.

Finally, teacher judgements may be influenced by a variety of emotional responses to their work. Gaede (1978), for instance, found that teachers' assessment of their own knowledge increased

as they moved through their teacher education programs, but decreased substantially during their first year of teaching. Certainly these teachers did not suddenly know less once they entered their own classrooms, but just as certainly, they *felt* they knew less once they encountered the demands of real teaching.

Credibility of the Argument

The logic of ask-the-teacher studies goes something like this: If teachers who choose to respond to the survey claim they are competent in certain areas, or if they claim they have (or have not) learned something valuable from their teacher education programs, we can assume they are correct and that their estimates of the contributions of teacher education are also correct. Since there is no direct measure of teachers' knowledge or skill, the burden of the argument falls entirely on the teachers' judgments.

Moreover, these studies almost never include comparison groups. Each study uses a unique survey instrument on a particular group of teachers who graduated from a particular institution. Consequently, it is extremely difficult to compare teacher judgments across studies—to say, for instance, that teachers from Program A felt the program had more benefit than teachers from Program B attributed to their program. An interesting effort designed to correct for this problem is currently under way at Ohio State University (Loadman and Gustafson, 1990), where a group of institutions have agreed to use a common survey instrument for their normal graduate follow-up studies. Once a sufficient number of institutions have conducted surveys with this instrument, it may be possible to draw some simple contrasts among respondents from different institutions or different types of institutions.

Finally, none of the studies take into account the teaching context. Some teaching situations are far more challenging than others; some provide less assistance to new teachers than others; and some provide considerably different expectations of teachers than their programs may have been striving for. To the extent that any of these contextual differences might influence teacher judgments, the findings are even more difficult to interpret.

Thus, to make any sense of these data, we have to assume that (a) teachers use the same criteria to judge themselves and their programs as teacher educators, policymakers, or educational researchers would use; (b) teachers' assessments of their own knowledge and skills are valid; and (c) the context in which teachers are teaching has no bearing on their assessments of themselves or their teacher education programs. And even after making these assumptions, we don't know what to make of the ratings that we see, for we have no comparison against which to gauge them.

Overall, then, ask-the-teacher studies have only limited utility. Although the aspect of teacher education they examine is central to reform efforts, in that they address the specific contents

of teacher education programs, each study is limited to a particular institution, so that generalizable conclusions are hard to draw. Even more important, though, is that the outcomes on which they focus are so seriously limited that the credibility of their argument is also jeopardized.

Experiments in Teacher Education

The fourth way to think about whether teacher education makes a difference is to test experimentally the contributions of teacher education. Researchers using this genre contrast particular approaches to teacher education and document changes in teacher candidates who are exposed to teaching. Much of this research was spawned during the era when microteaching was a dominant proposal for reforming teacher education. Researchers interested in microteaching, or in other specific aspects of teacher education, contrast two or more of these approaches in an effort to discern the relative merits of each.

Experiments avoid several of the limitations that the first three strategies have. They always contrast two or more clearly defined program variations, and they often include an assessment of the teachers' knowledge or skill prior to their participation in the study as well as after the study. And they usually directly assess the outcome of interest, rather than asking teachers to judge their own progress. In addition, they often randomly assign teacher candidates to the two or more program variations they are testing, to further ensure that groups receiving different variations do not differ in their motivations prior to participating in the study. These features give researchers a tremendous advantage, for they can not only tell us how their teacher candidates differed following exposure to different program variations, they can also tell us what these candidates were like before they participated.

Copeland's (1975) study of the relationship between microteaching and student teaching is a good example of microteaching experiments. In this study, Copeland first sorted students into two groups, one of which received microteaching training in the skill of "asking probing questions." He then observed a number of cooperating teachers and divided them into two groups, depending on the extent to which they tended to ask probing questions in their own teaching. Finally, he gave half of each group of cooperating teachers training in the supervision of student teachers. With these groups of cooperating teachers in place, he was able to assign his two groups of student teachers across the four groups of cooperating teachers, and to look at the combined effects of microteaching training with or without a cooperating teacher who demonstrated probing questions and with or without a cooperating teacher who had been trained in the supervision of student teachers. Copeland found that microteaching alone did not increase the likelihood that students would ask probing questions during student teaching, but neither did either of the other two treatments, either alone or together. However, the combination of all three forms of assistance did make a difference.

A more recent example of experiments is a series of studies reported at the annual meeting of the American Educational Research Association by Nancy Winitzky and Richard Arends (1989). These researchers first contrasted visits to exemplary classrooms with observations of videotape and found both to be equally effective in helping teachers use cooperative grouping in their own microteaching. In a second study, they contrasted two methods of developing novices' intellectual schemata regarding cooperative grouping; and in the third, they contrasted learning in the exemplary classrooms with learning via microteaching and found them to be equally effective. Like many such studies, these studies did not follow the students into their own student teaching experiences to see the extent to which they carried their new knowledge into their own teaching practice.

In his review of literature on laboratory experiences in teacher education, Copeland (1982) found that experiments tended to focus on four main aspects of microteaching: (a) the models used to train students, (b) whether novices teach real students or their own peers, (c) the type of feedback given after microteaching, and (d) the type of supervision provided. Each of these variations has been found to make a difference in some aspect of learning. More importantly, Copeland found that microteaching in general did improve the initial acquisition of teaching skills, but that the evidence was less than clear regarding the extent to which teacher candidates continued to use their new skills when they were teaching in real classrooms.

Now let's consider my three questions.

Aspects of Teacher Education Examined

More than any of the other studies, these studies tend to focus on highly definable and highly relevant aspects of teacher education. Researchers who conduct experiments are not interested in the amount of additional courses teachers take, as open-search researchers are, nor are they interested in whole certification programs, as are those researchers who compare arts and sciences graduates with teacher education graduates. Nor are they interested in teachers' retrospective judgments, as ask-the-teacher survey researchers are. They are interested in particular segments of teacher education and in rather fine-grained variations in strategies used within these segments of teacher education. On the surface, then, studies in this genre seem to be especially relevant to those who want to improve teacher education.

On the other hand, many of these studies suffer because they are *too* short in duration. They may contrast relatively small program units—three weeks of Approach A versus three weeks of Approach B or even three hours of A versus three hours of B. They do this, of course, in part because smaller units are easier to manage. But it is not clear that evidence of effectiveness within such small units can be used to make larger scale changes in the structure of teacher education programs.

Outcomes

With respect to outcomes, most of these researchers evaluate teacher candidates' abilities to perform the discrete skills for which they have been trained. They look, for instance, at candidates' questioning skills or at their skill in responding to student disruptions. The outcomes assessed are, by definition, directly relevant to teacher education, since they are selected specifically to reflect the program goals. But they often are limited to immediate impact: They examine teacher behavior immediately after the teachers complete these alternative program approaches. We do not know whether the changes observed at that time will be sustained several months later. Especially troublesome is that we do not know whether these immediate effects will be demonstrated once the teachers are teaching in their own classrooms. And it is their eventual classroom practice, after all, that we ultimately want to influence.

Yet another problem with these outcomes is that most of them are highly behavioristic. Researchers examine the extent to which teacher candidates have learned to employ a specific skill, but do not examine the extent to which candidates understand the point of using this skill or why it is valuable in teaching. Nor do they examine the teachers' affective response to the skill. If teachers learn to implement a skill on demand, but also learn to dislike the skill because of some other aspect of the experimental condition, we would not expect them to demonstrate the skill later on, when they are under no pressure to do so.

Credibility of the Argument

The logic of these studies is relatively simple and believable. If one program approach creates a greater increase in the target skill than others do, this approach has a greater impact than the others do. Because researchers have assessed their candidates' skills both before and after the candidates participated in their alternative approaches, and because they randomly assign candidates to the alternatives they have created, they can be more sure than other researchers that the differences they observe at the conclusion of the study do not reflect differences that were there in the first place.

Overall, then, these studies are more relevant to teacher educators in the aspects of teacher education they examine, more relevant in the outcomes they assess, and more powerful in their ability to draw unambiguous findings regarding the relative merits of one program approach over another. They could be strengthened a great deal by following their teacher candidates over a longer period of time, and by extending their outcomes beyond discrete behaviors.

Watch Teacher Candidates Change

The fifth way to think about whether or how teacher education makes a difference is to follow teacher candidates as they proceed through their college education, gathering data on them at several points along the way, to see whether and how their ideas about teaching change over time. Researchers working within this genre want to learn what students are like when they enter their programs, how they change over time in response to their programs, and what they are like when they finish. Like experiments, these studies offer us the advantage of being able to document change, so that if differences exist at the end of the study, we can interpret these differences relative to differences that may have existed at the outset. And like experiments, they sometimes enable us to look inside the black box, to see the details of the programs in which students participate and to see the interaction between the program and the students. Unlike experiments, though, these studies rarely allow us to compare students who participated in different kinds of programs. While we learn more about how students change as they encounter particular aspects of their programs, we cannot say with any confidence how they might have changed if they had participated in some other kind of program.

One of the earliest and best examples of this genre is Feiman-Nemser's and Buchmann's (1989) study of teacher candidates participating in two different teacher education programs. They followed six students participating in two preservice teacher education programs, interviewing them on several occasions about their understanding of what they were learning and about their views of teaching. They also observed the courses these students took. Through their descriptions of these students, they were able to demonstrate gradual shifts in views and to demonstrate ways in which the messages provided in these programs were occasionally misinterpreted by the candidates. The study demonstrates the importance of the teachers' entering assumptions and the ways in which they combine their own childhood experiences with the lessons they are being taught to form their own ideas about teaching and learning.

Another good illustration of this genre is Hollingsworth's (1989) study. She followed teacher candidates in a graduate program and through their teaching internships as well. Through her investigation, she was able to show not only the role that prior beliefs played in these teachers' learning but also how their university learning connected to their practice. She found that students' prior beliefs influenced their receptivity to the program and that they went through several distinct phases in their practice as they tried to accommodate what they had learned in the program to their classroom experiences.

And now to my three questions.

Aspects of Teacher Education Examined

The aspect of teacher education that these researchers tend to focus on is the particular patterns of courses that their sample students take. Even more particularly, they are often interested in courses *as they are perceived by the students themselves*. Instead of defining a program as consisting of a particular sequence of courses or other experiences, they define the program as the particular sequence of experiences that candidates respond to. A program brochure may claim, for instance, that Education 201 introduces students to findings from research on teaching. But the researcher who is documenting change in teacher candidates wants to know what Education 201 actually does. And in addition, he or she wants to know what Education 201 looks like to Student A, to Student B, to Student C, and so forth. Instead of allowing official program rhetoric to define the courses students take, they may actually attend courses with their sample students or ask students to describe what the faculty are telling them and what they make of that. Moreover, they are interested in how these courses accumulate over time to create particular changes in students. In this sense, the aspects of teacher education that they examine are highly relevant to teacher educators who are interested in restructuring their programs.

Outcomes

With respect to outcomes, these studies tend to be more interested in teachers' knowledge, beliefs, and attitudes than in their teaching skills, in part because they cannot really examine skills until teachers begin teaching and in part because these are the domains they expect to see changing as candidates participate in university courses. Researchers using this strategy take changes in candidates' beliefs and values, often as expressed by candidates in their own words, as their central outcome.

Credibility of the Argument

Many change studies are based on the assumption that teacher candidates enter their college programs with a set of initial beliefs that will influence their responses to the courses they take. As they participate in their courses, they respond by incorporating some new ideas but also by altering the messages they receive to make them more consistent with what they already believed. The influence of teacher education, or of college more generally, therefore, is not unidirectional. Instead, there is an interaction between students and their programs. Researchers who watch teachers change attempt to show how students who enter with different patterns of beliefs are influenced in different ways. They often gather extensive family and education background data on their students, and use these background data to interpret the changes they later observe.

The nature of this research is such that it is far more theory-dependent than research in the other genres. Since researchers are following students over time, since numerous possible changes can occur, and since these changes can be influenced by numerous possible student background characteristics as well as by numerous possible program characteristics, the quality of this research depends heavily on the quality of theory that guides data collection.

Among the five research genres reviewed here, this is the only one that assumes that the outcome of teacher education is a function not only of what the program teaches but also of what candidates believe when they enter their programs. Rather than looking to see whether candidates have acquired the particular knowledge or skills transmitted by a program, researchers in this genre are interested in the ways in which candidates' own beliefs interact with program messages to create a unique set of new ideas about teaching and learning.

One difficulty that some such studies encounter is a confusion between change due to student development and change due to program impact. The fact that students are changing and developing over time does not necessarily mean that these changes are a result of a program impact. College students are still in a highly formative stage in their lives and may be changing in several ways that have little to do with the particular courses or curricula they encounter as students. Thus, the credibility of change studies is highly dependent on the inclusion either of comparison groups, so that changes can be contrasted across program types, or of detailed background data and program data, so that the nature of the changes can be interpreted in light of these context variables. There are numerous ways in which these researchers try to separate out the effects of normal maturation from those of the program. Feiman-Nemser and Buchmann (1989), for instance, included students from two different programs and collected data on the programs as well as on the students. They increased the credibility of their argument by showing specific relationships between the ideas their students had and the ideas that were presented in their courses.

Another difficulty that can arise in change studies derives from the number of observations made on students. Students are often interviewed on numerous occasions, and it is highly likely that, over time, they learn what kind of responses their interviewers are looking for. Thus there is a chance that the researchers themselves are at least partly responsible for the changes they describe.

Overall, then, these studies focus on relevant aspects of teacher education— undergraduate programs and components within those programs—and on relevant outcomes—changes in knowledge and beliefs about teaching. The logic is also sound, provided that attention is given to sorting out natural maturation from program effects. While the findings are rich and informative and provide many insights into how college students interpret and respond to their undergraduate programs, they also are rather complex, leaving us with so many patterns of change that it may be difficult for us to gauge the extent to which any particular kind of change is occurring.

Conclusions

All five of these research genres are intended to document whether or how teacher education makes a difference. But even though they are designed to examine the same *general* question, they address it in quite different ways. Here are the questions they actually ask:

1. Do teachers who have taken more teacher education raise students' achievement more than teachers who have had less?

2. Do certified teachers teach differently than uncertified teachers?

3. Do teacher education graduates think they have the necessary knowledge and skill to teach, and do they think their teacher education courses helped them teach better?

4. Does one approach do better than another in helping teacher candidates learn specific skills?

5. How do the views of college students change as they participate in different kinds of undergraduate teacher education programs?

There are good reasons for this diversity of approaches, for the general question is large and complex. Each genre gives us a different perspective on the general issue. But since no single research project can reveal the full, complex, and amorphous picture, each is also necessarily limited. My aim here has been to demonstrate both the strengths and the limitations of these different genres, in the hope that researchers of all these persuasions might find ways to benefit from ideas in the other genres. Every researcher must make decisions that will limit the potential value of his or her study, and a better understanding of the trade-offs involved in these decisions can help researchers in their task. The three decisions I have focused on have to do with the aspects of teacher education that are examined, the outcomes that are examined, and the argument that connects these two together.

Choosing the Aspect of Teacher Education Examined

In reviewing these genres, it seems clear that some aspects of teacher education are more fruitful to examine than others. For instance, the first two genres I reviewed—the open searches and the comparisons of liberal arts graduates and teacher education graduates—have chosen to treat teacher education as a black box. Because they are studying teachers who have already completed their programs, they often know very little about the actual content and character of the teacher education programs themselves. So they treat teacher education as if it were a homogeneous, fixed

entity. This is a serious limitation for both policymakers and teacher educators, for research that treats teacher education as a black box tells us nothing about how the contents of that box might be rearranged or revised.

The remaining three genres all enable us to examine the contents of teacher education programs, but they do so in differing ways, and these differences also present different strengths and limitations. For instance, when a survey asks teachers to rate particular courses, we may still not know which section of the course each teacher took, or which faculty member taught the course to the teacher, or at what point in the teacher's curriculum sequence this course was taken. In contrast, many experiments and change studies are able to describe in detail the actual program components or courses that students take. Since these details are the aspects of teacher education that are most likely to matter, these are more fruitful aspects to study.

Choosing the Outcomes

It also seems clear from this review that there are numerous relevant outcomes that could be examined and that nearly all of these genres have examined a relatively narrow range of outcomes. Open searches limit their attention to student achievement test scores, as if these were the only outcomes teachers tried to influence, and experiments tend to focus on one or two specific skills. Ask-the-teacher studies limit their attention to teachers' judgments of their own capabilities, and we can never be sure what criteria teachers are using to judge their own knowledge and skills. Comparison studies and change studies both rely on broader ranges of outcomes; the former by observing real teaching in all of its complexity, and the latter by allowing teachers to express their ideas about a variety of topics. Still, neither of these genres incorporates the outcomes of the other.

Teachers may benefit from teacher education in many qualitatively different ways: They may acquire knowledge, alter their beliefs, gain skills, or develop new attitudes and dispositions. And all of these outcomes may be important to teaching practice. Moreover, in any given segment of teacher education, regardless of its primary intent, teachers may be influenced in more than one way. Even when the program is concentrating on skills, teachers will acquire some new knowledge and may change their beliefs or dispositions, particularly regarding the specific skills being taught. They may learn to perform a particular skill but may also learn to hate it and may even vow never to use it in their own practice. Any study that addresses only one of these outcomes is, therefore, automatically too narrow in its focus. If researchers measure only the particular skills they are aiming for, they won't know the full impact teacher education programs have had on their candidates.

Enhancing the Credibility of the Argument

Finally, an examination of these genres reveals the importance of designing research studies so that a credible argument can be made from the data. Three of the genres I have described here suffer in credibility because they examined teachers only *after* the teachers had completed their education, not before. In these genres—open searches, comparison studies, and ask-the-teacher studies—we have difficulty drawing inferences about whether or how teacher education has made a difference because we do not know how the various teachers in the study differed from one another *before* their college education and why different teachers chose the particular programs or courses that they did. The effects of program participation are confounded with the effects of self-selection into the programs.

The sad fact is that poorly designed studies are not merely *non*informative. Often, they are *mis*informative: By failing to consider what teacher candidates already knew prior to participating in teacher education, for instance, researchers may draw conclusions that either over- or underestimate the value of teacher education. To the extent that research genres misinform the field, they do a disservice to the field. They may mislead policymakers into adding or removing requirements erroneously from their programs or mislead teacher educators into over- or underusing particular program features. When researchers engage in open searches, they may erroneously conclude that recency of educational experiences enhances pupil achievement gains, when perhaps what really matters is the teacher's disposition to seek out ways to improve her own practice. When researchers choose to ask the teacher how what she learned, they may erroneously conclude that teachers did not learn much about classroom management, when perhaps teachers actually learned quite a bit about this, but are unaware of how little they knew before they studied teacher education.

The experiments and the studies of teacher change are least susceptible to this error and offer the most potentially credibly arguments about whether and how teacher education has made a difference. These two genres provide three advantages over the other three: Both allow us to see what teacher candidates are like before they participate in their programs, and both allow us to examine more closely the relationship between program character and content, on one hand, and outcomes on the other. Finally, both entail some theoretical work along with the empirical work, a feature that increases the likelihood that, over the long run, research findings will accumulate into a more meaningful body of knowledge about teacher education.

Notice, too, that these two genres rest on quite different assumptions about *how* teacher education is likely to make its difference. Experiments tend to focus more on program strategies, while change studies tend to focus more on program content. Experiments tend to define outcomes in terms of discrete, predefined skills, whereas change studies tend look for evidence of altered beliefs. Experiments tend to assume that program influences are unidirectional, whereas change

studies tend to assume that programs interact with candidates' entering ideas to produce new ideas about teaching. Experiments often derive from a behaviorist framework, whereas change studies often derive from a cognitive-constructivist framework. Thus, these two strategies are based on substantially different assumptions about what teachers need to learn and on substantially different assumptions about the relationship between programs and teacher candidates.

Yet they share several features that are important to those who want to reform or restructure teacher education programs, for both enable us to learn more about the details of how candidates respond and change as they participate in particular aspects of teacher education. These advantages suggest that we might move much further in our efforts to understand teacher education and how it works if we were to increase the effort we invest in experiments and in studies of teacher change.

From all of these research genres, then, we learn something about whether and how teacher education makes a difference. But we also learn, by examining the genres themselves, that the way a researcher poses his or her research question constrains what can be learned from the study. And we recognize that such constraining decisions are necessary, for the enterprise of teacher education is too large, complicated, and amorphous to be succumb to an all-encompassing study. The challenge facing researchers in teacher education is to maximize the potential of their studies by assuring that the aspects of teacher education they study are meaningful and relevant to teacher educators who want to use research to improve their programs, that the outcomes they examine are sufficient, and that the evidence they gather will enable them to develop reasonable arguments about whether and how teacher education has made a difference.

References

Adams, R. D., and Craig, J. R. (1983). A status report of teacher education program evaluation. *Journal of Teacher Education, 34*(2), 33-36.

American Association of Colleges of Teacher Education. (1987). *Teaching teachers: Facts and figures*. Washington, DC: Author.

American Association of Colleges of Teacher Education. (1989). *RATE III: Teaching teachers: Facts and figures*. Washington, DC: Author.

Arch, E. C. (1989, March). *Comparison of student attainment of teaching competencies in traditional preservice and fifth-year master of arts in teaching programs*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Beery, J. R. (1960). *Professional preparation and effectiveness of beginning teachers*. Coral Gables, FL: Graphic Arts Press.

Begle, E. G., and Geeslin, W. (1972). *Teacher effectiveness in mathematics instruction* (National Longitudinal Study of Mathematical Abilities Reports: No. 28). Washington, DC: Mathematical Association of America and National Council of Teachers of Mathematics.

Brown, D., Edington, E., Spencer, D. A., and Tinafero, J. (1989). A comparison of alternative certification, traditionally trained, and emergency permit teachers. *Teacher Education and Practice, 5*(2), 21-23.

Carpenter, T. P., Fennema, E., Peterson, P. L., Chang, C., and Loef, M. (1989). Using knowledge of children's mathematics thinking in classroom teaching: An experimental study. *American Educational Research Journal, 26*, 499-531.

Clark, D. C., Smith, R. B., Newby, T. J., and Cook, V. A. (1985). Perceived origins of teaching behaviors. *Journal of Teacher Education, 36*(6), 49-53.

Coladarci, T., and Gage, N. L. (1984). Effects of a minimal intervention on teacher behavior and student achievement. *American Educational Research Journal, 21*, 539-555.

Coleman, J., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., and York, R. L. (1966). *Equality of educational opportunity*. Washington, DC: U.S. Office of Education.

Copeland, W. D. (1975). The relationship between micro-teaching and student teacher classroom performance. *Journal of Educational Research, 68*, 289-293.

Copeland, W. D. (1982). Laboratory experiences in teacher education. In H. E. Mitzel (Ed.), *Encyclopedia of educational research* (Vol. 2, 5th ed., pp. 1008-1019). New York: Free Press.

Cornett, L. (1984). *A comparison of teacher certification test scores and performance evaluations of graduates in teacher education and liberal arts and sciences in three southern states*. Atlanta, GA: Southern Regional Education Board.

Council of Chief State School Officers. (1988). *State education indicators, 1988*. Washington, DC: Author.

Devalt, M., and Ball, D. W. (1987). Some effects of training on the competence of beginning teachers. *Journal of Educational Research, 80*, 343-347.

Drummond, R. J. (1976). *Follow up of 1970-76 College of Education Graduates*. University of Maine, Orono. (ERIC Document Reproduction Service No. ED 141 306)

Feiman-Nemser, S., and Buchmann, M. (1989). Describing teacher education: A framework and illustrative findings from a longitudinal study of six students. *Elementary School Journal, 89*, 365-377.

Gaede, O. F. (1978). Reality shock: A problem among first-year teachers. *The Clearinghouse, 51*, 405-409.

Gliessman, D. H. (1987). Changing complex teaching skills. *Journal of Education for Teaching, 13*, 267-275.

Good, T. L., and Brophy, J. E. (1986). School effects. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 570-603). New York: Macmillan.

Good, T., Grouws, D., and Ebmeier, H. (1983). *Active mathematics teaching*. New York: Longman.

Goodman, J. (1986). What students learn from early field experiences: A case study and critical analysis. *Journal of Teacher Education, 36*(6), 42-48.

Griffin, G., and Barnes, S. (1986). Using research findings to change school and classroom practices: Results of an experimental study. *American Educational Research Journal, 23,* 572-586.

Hanushek, E. A. (1971). Teacher characteristics and gains in student achievement: Estimation using micro data. *American Economic Review, 61*, 280-288.

Hanushek, E. A. (1972). *Education and Race: An analysis of the educational production process*. Lexington, MA: D. C. Heath.

Hanushek, E. A. (1989). The impact of differential expenditures on school performance. *Educational Researcher, 18*(3), 45-51; 62.

Henry, M. (1986). Strengths and needs of first-year teachers. *Teacher Educator*, *22*(2), 10-18.

Hodges, C. (1982). Implementing methods: If you can't blame the cooperating teacher who can you blame? *Journal of Teacher Education*, *33*(6), 25-29.

Hollingsworth, S. (1989). Prior beliefs and cognitive change in learning to teach. *American Educational Research Journal, 26*, 160-189.

Howey, K. R., and Zimpher, N. L. (1989). Preservice teacher educators' role in programs for beginning teachers. *Elementary School Journal, 89*, 451-471.

Howey, K. R., and Zimpher, N. L. (1990). Professors and deans of education. In R. W. Houston, M. Haberman, and J. Sikula (Eds.), *Handbook of research on teacher education* (pp. 349-370). New York: Macmillan.

Kluender, M. M. (1984). Teacher education programs in the 1980's: Some selected characteristics. *Journal of Teacher Education*, *35*(4), 33-35.

Loadman, W., and Gustafson, G. L. (1990). *National database for teacher education program follow-up*. Columbus, OH: Ohio State University College of Education.

Lupone, O. J. (1961). A comparison of provisionally certified and permanently certified elementary teachers in selected school districts in New York State. *The Journal of Educational Research*, *55*(2), 53-63.

McIntyre, D. J., and Killian, J. E. (1986). Students' interactions with pupils and cooperating teachers in early field experiences. *Teacher Educator*, *22*(2), 2-9.

McIntyre, D. J., and Killian, J. E. (1987). The influence of supervisory training for cooperating teachers on preservice teacher development during early field experiences. *Journal of Educational Research*, *80*, 277-282.

Medley, D. (1982). *Teacher Competency Testing and the Teacher Educator*. Charlottesville, VA: University of Virginia.

Murnane, R. J., and Phillips, B. R. (1981). What do effective teachers of inner-city children have in common? *Social Science Research, 10*, 83-100.

National Center for Education Statistics. (1989). *Digest of education statistics 1989*. Washington, DC: Author.

National Center for Research on Teacher Education. (1991). *Initial findings from the Teacher Education and Learning to Teach (TELT) Study*. Unpublished paper, Michigan State University, National Center for Research on Teacher Education, East Lansing.

Pigge, F. L. (1978). Teacher competencies: Need, proficiency, and where proficiency was developed. *Journal of Teacher Education, 29*(4), 70-76.

Reed, H. B. (1975). *Evaluation of alternative approaches to teacher preparation*. University of Massachusetts. (ERIC Document Reproduction Service No. ED 147 335)

Schmelter, R. C. (n.d.) *A Follow up study of mathematics education majors from the University of Wisconsin-Oshkosh, 1985*. (ERIC Document Reproduction Service No. ED 265 034)

Silvernail, D. L., and Costello, M. H. (1983). The impact of student teaching and internship programs on preservice teachers' pupil control perspectives, anxiety levels, and teaching concerns. *Journal of Teacher Education, 36*(4), 32-36.

Skipper, C. E., and Quantz, R. (1987). Changes in educational attitudes of education and arts and sciences students during four years of college. *Journal of Teacher Education, 38*(3), 39-44.

Smylie, M. A. (1989). Teachers' views of the effectiveness of sources of learning to teach. *Elementary School Journal, 89*, 543-558.

Strang, H. R., Badt, K. S., and Kauffman, J. M. (1987). Microcomputer-based simulations for training fundamental teaching skills. *Journal of Teacher Education, 38*(1), 20-26.

Tabachnik, B. R., and Zeichner, K. M. (1984). The impact of student teaching experience on the development of teacher perspectives. *Journal of Teacher Education, 35*(6), 28-36.

Veenman, S. (1984). Perceived problems of beginning teachers. *Review of Educational Research, 54*, 143-178.

Warren, R. D., Dilts, H. E., Thompson, A. D., and Blaustein, M. H. (1982). *Follow-up study of first year Iowa State Graduates. General Frequency Report*. Ames: Iowa State University. (ERIC Document Reproduction Service No. ED 270 436)

Winitzky, N. E., and Arends, R. (1989, March). *Moving research into practice: The effects of various forms of training and clinical experiences on preservice students' knowledge, skills,*

and reflectiveness. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.