Contributions of Qualitative Research to Research on Teacher Qualifications
Mary M. Kennedy
EDUCATIONAL EVALUATION AND POLICY ANALYSIS 2008; 30; 344
DOI: 10.3102/0162373708326031

The online version of this article can be found at:
http://epa.sagepub.com/cgi/content/abstract/30/4/344
Contributions of Qualitative Research to Research on Teacher Qualifications

Mary M. Kennedy
Michigan State University

The influence of teachers' qualifications on their teaching practice has been subject to debate. Literature reviews do not settle these debates, partly because the literature is uneven and partly because reviews capture only narrow slices of literature. In particular, many reviews eliminate qualitative studies. Yet without examining qualitative evidence, variations in quantitative findings are difficult to interpret, disappointing findings are difficult to understand, and plausible explanations of patterns are in short supply. The present review focuses on qualitative studies and compares their findings with those from quantitative literature. The author finds that the qualitative literature agrees with quantitative literature in its inability to distinguish between teachers with different types of certificates or different teacher education backgrounds. On the other side, the author finds more evidence of benefit from content knowledge than quantitative studies have typically found. The qualitative studies also reveal competing influences and offer hypotheses about why the outcomes look the way they do.

Keywords: teacher qualifications, qualitative methods, teacher education

The volume of research on teachers' qualifications has grown to several hundred studies, but it has not settled arguments about the merits of teacher education programs. One reason for this stalemate is that the question is difficult to address, and virtually all studies are susceptible to confounding variables. For instance, teachers self-select their educational programs; consequently, their own values and predispositions are confounded with their credentials. And once certified, they engage in nonrandom job-seeking practices, while districts engage in nonrandom hiring practices, so the resulting pattern of job placements confounds teachers’ educational backgrounds, credentials, attitudes, and predispositions with the communities they serve (Boyd, Lankford, Loeb, & Wykoff, 2003; Lankford, Loeb, & Wykoff, 2002; Strauss, 1999). Kennedy, Ahn, and Choi (2008) call the resulting allocation of teachers to schools a system of affinity assignments, meaning that teachers’ social backgrounds and qualifications often match the social backgrounds and “qualifications” of their students.

The problem is compounded when literature reviews use different inclusion rules. Some reviewers (e.g., R. Greenwald, Hedges, & Laine, 1996; Wayne & Youngs, 2003; Wilson, 2005) call the resulting allocation of teachers to schools a system of affinity assignments, meaning that teachers’ social backgrounds and qualifications often match the social backgrounds and “qualifications” of their students.

The author acknowledges financial support from the U.S. Department of Education’s Office of Educational Research and Improvement, now Institute for Educational Sciences; by the National Science Foundation’s Division of Research, Evaluation, and Communication; and by the National Science Foundation’s Math and Science Partnerships. However, the opinions expressed here are the author’s, and no institutional endorsement should be inferred.
Research on Teacher Qualifications

Floden, and Ferrini-Mundy, 2001) do not reach beyond journal articles, whereas others cast a broader net. Some consider only studies that use student achievement as their outcome measure (e.g., R. Greenwald et al., 1996; Wayne & Youngs, 2003), whereas others include an array of indicators of quality teaching (Rice, 2003; Wilson et al., 2001). Some (e.g., Kennedy et al., 2008) require that outcomes be measured after teachers have finished their education and are employed as teachers of record, whereas others (e.g., Wilson et al., 2001) include studies whose outcomes were measured while teachers were still participating in student teaching. Some are unclear about their inclusion rules (e.g., Walsh, 2001). Finally, each review also imposes its own criteria for study quality. The number of studies included in reviews of literature on teacher qualifications ranges from 21 in Wayne and Youngs’s (2003) review to 200 in Walsh’s (2001) review.

In part because different readers of this literature focus on different subsets within it, we also can see very different conclusions arising as different authors perceive the available evidence in very different ways. Here are some illustrative quotes:

More than 200 studies have found that teachers who have more background in their content areas and have greater knowledge of teaching and learning are more highly rated and more successful with students in fields ranging from early childhood and elementary education to mathematics, science and vocational education. (Darling-Hammond & McLaughlin, 1999, pp. 377–378)

I would conclude, if anything, that there is little support for the impact of teachers’ subject matter knowledge on student learning. (Friedman, 2000, p. 20)

Contrary to conventional wisdom, mathematics and science students who have teachers with emergency credentials do no worse than students whose teachers have standard teaching credentials, all else being equal. (Goldhaber & Brewer 1999, p. 97)

The evidence indicates that neither an extreme centralized bureaucratization nor a complete deregulation of teacher requirements is a wise approach for improving teacher quality. (Rice, 2003, p. vi)

These quotes reveal that the relationship between teacher qualifications and teaching quality—a relationship that should be self-evident—is not self-evident at all. Moreover, the field lacks coherent theoretical arguments about how qualifications are expected to influence teaching practice. This is an area, therefore, where qualitative research could make a strong contribution, for qualitative research is more able to delve into the mechanisms and processes by which these qualifications actually influence teaching practice.

My contribution to this debate is to add qualitative studies to the corpus of literature under review. Because these studies have rarely been considered in reviews, I also offer a framework for evaluating the validity of their causal claims. And to gauge the import of their contribution, I contrast their conclusions with those from other literature reviews, and I ask whether they help fill the theoretical gap regarding how knowledge and educational backgrounds influence teaching practice.

Most reviews exclude qualitative research, either because its outcomes are not quantifiable or because of a perception that qualitative studies are useful for learning about social meanings but not about causal influences. Yet many qualitative studies do examine the influence of policies or programs on teaching practice, and they can help us understand these influences in a way that quantitative studies cannot. Maxwell (2004a, 2004b) has argued that qualitative studies address causation not by establishing patterns of regularities, as quantitative studies do, but by revealing the mechanisms and processes of influence. In the case of teaching, qualitative studies can help us see how a teacher’s knowledge either misdirects practice or enhances it and see where and how educational backgrounds have their influences, if they do. Qualitative research is also the preferred method of teacher educators, as evidenced by the large fraction of qualitative studies in journals such as the Journal of Teacher Education, so when reviewers exclude this literature, they are excluding the bulk of work that has been done by teacher educators themselves.

Qualitative research has been a staple of social sciences for some time but ascended in educational research in the 1980s. Since then, researchers have struggled to articulate the differences between quantitative and qualitative research and to articulate standards of validity for qualitative research. It has often been referred to
not merely as an alternative method but as an alternative paradigm, with different assumptions, methods, rationale, and purposes. Some reviewers (e.g., Wilson et al., 2001) label this work “interpretive,” suggesting that the nature of one’s data automatically implies an alternative epistemology. Others (e.g., Becker, 1996) insist that although qualitative and quantitative approaches are different, their epistemologies are not. Maxwell (2004a, 2004b) argued that the two approaches are subject to many of the same concerns regarding validity but that the former relies on variance and regularities to establish causality, whereas the latter relies on revealing the causal mechanisms themselves. Donmoyer (2001) argues that there is no single way to characterize qualitative studies, for they vary substantially in their aims. He suggests that they be defined by their purpose rather than by their epistemologies or paradigms. Included in his list of purposes is one called “truth-seeking.” Truth-seeking qualitative studies differ from those whose purpose is, say, to create a personal interpretation and those whose purpose is social change. The fact that meta-analysts are now seeking approaches to incorporating qualitative studies into their work (see, e.g., Hundersmarck, 2004; Au, 2007) suggests a growing interest in the truth-seeking potential of qualitative studies.

The central question guiding the present review is whether and how truth-seeking qualitative studies can amend, clarify, or otherwise add to the knowledge gained from the larger quantitative body of literature regarding the influence of teachers’ qualifications on the quality of their teaching practices. All of the studies reviewed here examine the influence of teachers’ knowledge or educational backgrounds on the quality of their teaching practice, and all seemed to expect their findings to be generally relevant. In this sense, they fit Donmoyer’s (2001) category of truth-seeking studies.

The article has three main sections. In the first, I describe the literature search process. In the second, I examine methodological issues associated with truth-seeking qualitative studies and offer some criteria for evaluating their validity. In the third, I review the findings from these studies. I organize the findings around different kinds of qualifications.

### Search Procedures

This review is based on literature in the Teacher Qualifications and the Quality of Teaching (TQQT) database. This database contains a collection of research articles, reports, dissertations, conference presentations, and books that examine the relationship between at least one teacher qualification and at least one indicator of the quality of teaching, with qualifications defined as aspects of education or certification that have been the target of education policy. Literature in ERIC, PsycInfo, and EconLit was searched using more than 20 terms to characterize teachers’ educational backgrounds: teacher knowledge, test scores, preparation, education, courses taken, and so forth. The search also included more than 20 terms to characterize evidence of teaching quality: authentic practice, quality of practice, student achievement, classroom practice, classroom activities, and so forth. Studies had to be published after 1960 and had to investigate K–12 teachers of record in the United States. Studies of student teachers were eliminated, as were those that placed teachers in artificial settings or in simulations. This last requirement eliminated a number of small-scale experimental studies.

Each study also had to establish a link between the qualification and the quality of practice, rather than simply describing both and asserting that they were related. Here are two examples of qualitative studies that were rejected for failure to establish a link. In one case (Pool, Ellett, Schiavone, & Carey-Lewis, 2001), the authors purported to test the link between National Board Certification and teaching quality. They drew a sample of six board-certified teachers, interviewed them, and observed their practices. They found that the teachers’ practices were quite variable and argued that this variability constituted evidence that board certification was not associated with teaching quality. We rejected the argument that variability within a group was, by itself, evidence that group was not distinguished relative to the population as a whole. In another study, Wozniak (1990) set out to learn the backgrounds of a sample of K–12 art teachers who had been
identified as outstanding. The author then reported common themes and argued that the characteristics she found must have been responsible for the quality of teachers’ practice. Nothing in the study showed readers how these teacher characteristics actually fostered teaching quality, nor did the author examine a comparison group of less outstanding teachers to see if they lacked these qualifications or had them. Notice that in both of these cases, studies were rejected because of the logic of their claim, not the quality of the evidence.

The results of this search-and-screen process was a database of more than 450 study reports, including 23 studies describing findings that were primarily qualitative but that spoke to the causal question of whether or how teachers’ knowledge or educational backgrounds influenced the quality of their teaching practice. Below, I discuss methodological issues uniquely associated with these 23 qualitative studies.

Defining Validity in Truth-Seeking Qualitative Research

A great deal of ink has been spilled in the cause of defining validity for qualitative research, and my aim is not to review all of that here. Instead, I concern myself with validity issues associated with a narrow slice of qualitative research that is truth seeking in its focus and that specifically addresses the influence of teachers’ knowledge and educational backgrounds on the quality of their teaching practices. For this group of studies, criteria for validity are very similar to those faced by quantitative researchers: Studies must show evidence that an influence actually occurred or failed to occur, they must persuade readers that the observed teachers are not idiosyncratic in some way, and they must acknowledge, and try to eliminate, any alternative explanations of their findings. I address each of these issues below, beginning with the alternative explanations that must be overcome and then reviewing study designs and analytic strategies that these researchers employed. The issues I enumerate below are not intended to apply to all truth-seeking qualitative research but to the specific studies in this database.

Alternative Explanations for Qualitative Findings

Whether quantitative or qualitative, research on teacher qualifications is virtually all non-experimental. No researcher has randomly assigned prospective teachers to different kinds of certificates, majors, grade point averages, or test scores nor assigned them to different school districts or schools. Therefore, all of these studies must be examined for alternative explanations for observed patterns in outcomes. These alternatives threaten the validity of the inference that the qualification of interest was responsible for observed outcomes. Here is a list of alternative hypotheses that most often presented themselves in this set of 23 studies, listed in order of how frequently they appeared.

Accommodation bias. One explanation for observed practice is that teachers are trying to accommodate the researcher’s interests and values. Because qualitative researchers have numerous interactions with teachers, they have many opportunities to share their own values and beliefs and consequently to “tell” teachers what kind of practices they want to see or what values and beliefs they want to hear about. The problem is especially severe when the researcher is a teacher educator studying teachers who were his or her students a few years earlier. In these cases, teachers would be well aware of pedagogical theories and practices the researcher advocated. I do not mean to suggest that teachers would invent completely phony personalities; they could simply emphasize some aspects of their work more than others, use the researchers’ language to describe their views and practices, or make an effort to generate examples of how a program had influenced them, simply to be helpful. These adjustments are a standard part of social practice, but qualitative researchers need to be alert to this possibility and develop data collection procedures that can correct for this potential bias. Direct observation of practice, rather than reliance on interviews, is a strong corrective for accommodation bias, as teachers would be unlikely to generate a practice that they did not normally produce.
Confirmation bias. Observers may be more inclined to notice events that fit their own a priori hypotheses than things that do not. For example, a researcher who believes a traditional teacher preparation program is superior to an alternative program may notice more problems when observing teachers who were prepared in an alternative program. Confirmation bias is a widely recognized human disposition and can occur in research as well as in everyday life (A. G. Greenwald, Pratkanis, Leippe, & Baumgardner, 1986). It can be controlled by ensuring that observers do not know which teachers come from which educational backgrounds (a procedure ironically called “blind observation”), though in this collection of studies, I found only two in which observers were blind to the qualifications of the teachers they were observing. There are also techniques available to researchers who videotape lessons and then assign “blind” observers to document events. Maxwell (2004b) offers other strategies that can help here as well.

Hindsight. One thing that may be unique to this body of literature is an evaluative tone in depictions of classroom practice. Researchers hold idealized visions of what good teaching looks like and often have videotapes that can be viewed and re-viewed, so their critique is based on repeated opportunities to reflect on events and on how things could have or should have been done. But the teacher’s practice was generated spontaneously and did not have the benefit of reflective hindsight, so the interpretation of practice may be overly idealist and unfair to the teacher. One possible antidote for hindsight is to accompany critiques of practice with evidence of the teacher’s own rationale for what occurred and the teacher’s own critique of what occurred.

Reactivity. Sometimes the data collection process itself can influence study outcomes. For instance, if a researcher is examining the relationship between teachers’ knowledge of, say, cultural differences among students and their classroom practices, he or she might begin by asking teachers about the issue, thus increasing the teachers’ awareness of it and thereby altering their subsequent practice. Reactivity can be reduced by using more indirect interview techniques that pursue particular topics without overtly announcing their salience to interviewees.

Misspecified influences. Researchers interested in the influence of a particular teacher education program sometimes document program content after the teachers have graduated, so the program the researchers saw at the time of the study might not be the program that their focal teachers saw when they were students. The best way to prevent this is to document programs at the time prospective teachers are experiencing them rather than later on.

Confounded contexts. Because qualitative researchers tend to focus on only a handful of cases, differences among teaching practices may be influenced by idiosyncratic aspects of local contexts rather than by the teachers’ degree or other qualification. Perhaps one teacher teaches college preparatory classes whereas the other teaches remedial classes, or perhaps one is in a more supportive school or has more planning time than the other. It is possible that these variations, rather than a difference in qualifications, are responsible for differences in observed practice. Quantitative researchers, with their larger samples, hope that these variations will cancel each other out so that the overall influence of a particular qualification can be found above the din. A benefit of qualitative research is that it can incorporate local contextual influences into the analysis, but researchers are very uneven in the attention they give to these issues. In small samples, these aspects of context need to be incorporated into the researchers’ causal accounts.

Establishing Influence in the Face of Alternative Explanations

Studies in this review were all interested in how a particular qualification—an educational background or a body of content knowledge— influenced teaching practice. Thus a central task for researchers is to establish that such an influence did exist and that the appearance of an influence cannot be explained by reference to some other explanation, such as one of those listed above. The studies in this review used a relatively small set of strategies for establishing
that a qualification had some influence on practice, often relying on one or more of the following strategies: (a) between-teacher comparisons; (b) within-teacher comparisons; (c) longitudinal studies, typically following people from their teacher-education programs into their practice; (d) teacher attributions of influence; and (e) detailed descriptions that demonstrate the influence itself.

Between-teacher comparisons. About half the studies engaged in between-teacher comparisons, using a comparison-group logic that was analogous to quantitative group comparisons. For example, Powell (1997) contrasts two secondary science teachers, one who had not majored in the subject she was teaching and one who had an advanced degree in the subject plus several years of applied experience in the field. The two teachers participated in the same teacher education program and, upon graduation, taught the same secondary science courses.

Within-teacher comparisons. Two studies in this group relied on within-teacher comparisons. They contrasted a single teacher’s practice across two different subjects. If the research question has to do with something relevant to the content taught, such as the teachers’ subject matter knowledge, this can be a strong research design. Because the same teacher is teaching both subjects, a wide range of other aspects of the teacher’s background, attitudes and values, students, and teaching situation are held constant, so observed differences are easier to interpret.

Longitudinal and content-matching studies. Six studies employed a technique of following teachers from their undergraduate experiences into their professional teaching assignments to see if the ideas taught by the programs appeared in their graduates’ practices or rhetoric. There is a greater burden on these researchers to explain how they selected their samples of teachers, for the potential for bias is large. These designs also are threatened by the potential for accommodation bias, because teachers are aware of program goals and values.

Teacher attribution of learning. An obvious strategy for linking teachers’ educational backgrounds to their current teaching practice is to ask them where they got their ideas. No researcher used this as the only method for establishing influence, but several used this in conjunction with other strategies. The strategy is somewhat more risky if teachers know the purpose of the study and want to please the researcher. However, a good researcher will not ask, “Did you get that idea in our XYZ program?” but instead ask, “Where did you get that idea?”

Detailed descriptions of the influence itself. Although detail is a widely advertised strength of qualitative research, it is not widely practiced. I found only two studies that provided voluminous and detailed descriptions of observed practices to establish their causal links (Smith, 1989; Stein, Baxter, & Leinhardt, 1990). Both were case studies of individual teachers, and both provided numerous detailed examples. Such descriptions can overcome the lack of a comparison as well as many other alternative explanations.

Establishing Representation
With a Small Sample

The third challenge facing all researchers is to be clear about how their samples are selected and what each case is a case of. Teachers may be selected because they participated in different programs, participated in the same program but responded differently, entered a program with different prior beliefs and values, or took teaching positions in different contexts or for some other reason. The rationale for both initial selection and for contrasts is a central part of arguments about what was learned from the study. Some authors gave details about the particular teacher or teachers they studied, why they selected them and why they saw these particular teachers as newsworthy; others gave no explanation for their samples.

Critics of qualitative research often argue that these small samples lack value because they automatically deny any possibility of representativeness. And some qualitative researchers apparently agree, for instead of making a case for the representativeness of their samples, they rely instead on Yin’s (1989) argument that qualitative studies do not generalize to populations but only to theory. This argument does not
absolve researchers, however, for readers will necessarily ask, Who are these teachers and where did they come from? Readers need to know what the researchers saw as important in a prospective research subject, and they need to know the kind of teaching assignments, school environments, and students that sampled teachers contended with. This kind of information can help readers make their own judgments about the logic of any causal conclusions, about potential sources of bias in the study, and about the theoretical relevance of the findings as well.

In fact, the most widespread weakness in the studies reviewed here was a general failure to articulate how teachers were selected. Four studies gave no indication at all of how cases were selected, one based selection on geographic proximity, and another stated explicitly that the study capitalized on a serendipitous event. Many of the researchers already knew their sampled teachers because they had taught them, and yet they failed to say why they selected these graduates rather than some others.

### Study Findings

Findings are presented according to the particular type of qualification examined. I found studies examining the following: (a) influence of content knowledge that is stipulated on the basis of a credential, (b) influence of content knowledge as assessed by the researcher, (c) influence of particular teacher education courses, and (d) influence of broader patterns of courses or entire teacher education programs.

#### Influence of Stipulated Content Knowledge

Policy makers have access to a large volume of quantitative studies examining the relationship between teachers’ content knowledge and their students’ achievement. Content knowledge has been determined by college majors, degrees and certificates, test scores, and number of courses taken. Overall, this work has been disappointing in its failure to demonstrate the expected relationship. For example, one recent review (Ahn & Choi, 2004) synthesizing correlational research in mathematics education found an average correlation of only .11 between teachers’ tested content knowledge and their students’ achievement scores. Another review (Kennedy et al., 2008) focused on the number of mathematics courses teachers took in college but still found either very small or very inconsistent results across their studies. Many of these studies focused on knowledge of mathematics, where the lack of association between teachers’ content knowledge and their students’ achievement growth has been noticed since at least Begle (1972). Some mathematics educators (e.g., Ball, Hill, & Bass, 2005; Ball, Lubienski, & Mewborn, 2001) are now hypothesizing that this long history of mediocre findings might occur because teacher tests are not tapping the content knowledge that is most relevant to teaching. Studies examined in the next section consider a variety of conceptions of content knowledge.

In this section I review three studies that examine the role of content knowledge when the knowledge itself is not directly assessed but instead is stipulated by the presence of a degree or credential. The first study involves an author contrasting his own teaching in the field for which he felt prepared and in another field where he felt less qualified. The second study compared a teacher with professional work experience as a scientist with one who had no applied experience and who had majored in a different branch of science. The third compared two secondary English teachers, one of whom was certified to teach English, the other of whom was certified to teach a different subject. These studies are summarized in Table 1.

In the first study, Carlsen (1997) compared his own teaching in a subject he knew well (biology) with his teaching in a subject he knew less well (chemistry). I consider this to be stipulated knowledge because Carlsen did not provide any evidence of his knowledge other than to say that he had it. To see how his knowledge influenced his practice, he tape-recorded his lessons so that he could systematically examine the transcripts. He was particularly interested in the role that his questions played in controlling conversation and in reinforcing the teachers’ authority. He felt that an overreliance on questions could create an inquisition atmosphere in the classroom and that an overreliance on recall questions, without attention to rationale, reduced the authority of the content. In his comparison,
Carlsen saw that he used remarkably more questions when teaching chemistry than when teaching biology and, in particular, that he used more low-level, recitational questions. He also found that his chemistry lessons were much thinner in terms of revealing warrants for knowledge claims, so his chemistry students learned less about how facts are generated and evaluated than his biology students learned. He theorized that the difference between his questioning tendencies in chemistry versus biology occurred because he himself was more uncertain of the content and therefore relied on questions that tacitly discouraged, rather than encouraged, student discussion.

The second study (Powell, 1997) contrasted two novice teachers who graduated from the author’s own teacher education program. Both had completed their teaching degrees and had taken positions teaching earth sciences to lower-track high school students. They also shared many teaching values: They both espoused progressive pedagogies, both viewed science as a way of thinking and reasoning that students should experience firsthand, and both wanted to create a science curriculum that was embedded in activities rather than in the textbook. However, Jill’s content expertise was in biological sciences, not earth sciences, whereas Dan had a master’s degree in geology and had been a field hydrogeologist for 6 years before entering the teacher education program.

Powell found that both teachers had difficulty implementing their vision. Jill had difficulty because of her lack of knowledge of earth science. Dan, who had greater content knowledge and practical experience as well, did not know how to transform his deeper content knowledge into meaningful learning activities. So both teachers, despite their differences in content knowledge, ultimately compromised their plans to make science more active and meaningful to students and instead relied heavily on the textbook and on recitational teaching.

In the third study, Ringstaff and Sandholtz (2002) compared two secondary English teachers, one teaching out of field (Brian, certified in science) and one certified in English (David). Both graduated from the same university, participated in the same teacher preparation program, were 1st-year teachers, and received little support from their colleagues. The authors watched them teach *Cannery Row* and found that the presence of a teaching certificate in English did not promote better teaching. Brian, the science major, taught a low-track class but set high goals for his students, whereas David, the English major, taught a college preparation class but deemphasized literary analysis and focused instead on low-level knowledge. The authors conclude that the official degree is not a perfect indicator of content knowledge, nor of intellectual interest in the content, for Brian was an avid reader. Second, they note that participation in teacher education is not a perfect indicator. Both teachers attended the same teacher preparation program, but Brian was generally more able to manage his students and garner their interest in the content. Third, time available for planning could have been a complicating factor, because

---

**Table 1**

<table>
<thead>
<tr>
<th>Citation</th>
<th>n</th>
<th>Teaching responsibility</th>
<th>Qualification of interest</th>
<th>Evidence of quality of teaching practice</th>
<th>Strategy for establishing link</th>
<th>Link established?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carlsen (1997)</td>
<td>1</td>
<td>Secondary science</td>
<td>College major</td>
<td>Questioning strategies</td>
<td>Within-teacher comparison</td>
<td>Visible influence</td>
</tr>
<tr>
<td>Powell (1997)</td>
<td>2</td>
<td>Secondary science</td>
<td>Degree and experience</td>
<td>Broad approach to teaching</td>
<td>Between-teacher comparison</td>
<td>Weak influence</td>
</tr>
<tr>
<td>Ringstaff &amp; Sandholtz (2002)</td>
<td>2</td>
<td>Secondary English</td>
<td>In field vs. out of field</td>
<td>Broad approach to teaching</td>
<td>Between-teacher comparison</td>
<td>No visible influence but findings compromised by confounded context</td>
</tr>
</tbody>
</table>
David had four preparations per day whereas Brian had only two.

Influence of Assessed Content Knowledge

Eight studies directly assessed teachers’ content knowledge and then looked at how that was linked to their teaching practices. Their strategies were remarkably varied, and many incorporated aspects of quantitative research as well as qualitative research. They are summarized in Table 2. For ease of discussion, I sort them by grade span and content domain.

Elementary mathematics. Three studies examined the influence of teachers’ content knowledge on teaching practices in elementary mathematics. The first study examined 20 experienced teachers who had been in the same school for at least 3 years. Blasquez (1998) used both quantitative and qualitative methods to see the extent to which teachers’ understanding of mathematical concepts influenced their ability to use more constructive pedagogies. She used a set of structured tasks to assess their understanding of mathematical concepts and used a short questionnaire to learn their beliefs about procedural versus constructivist approaches to teaching. In her quantitative analysis, Blasquez found that teachers’ practices were more strongly associated with their pedagogical beliefs than with their mathematical knowledge. For the qualitative portion of the study, Blasquez selected three pairs of teachers matched in their beliefs about constructivist versus procedural instruction, but differing in their understanding of mathematical concepts, and then observed their lessons. Within each pair, Blasquez was able to see that the teacher with higher conceptual understanding was more likely to ask students to justify their ideas, more likely to extend or elaborate on student ideas, and more likely to monitor for understanding. So her qualitative analysis allowed her to see an influence that her quantitative work could not reveal. At the same time, however, the author was no longer “blind” for this portion of the study and knew which teacher belonged in each knowledge and belief category, so the things she noticed and wrote about could have been influenced by her own confirmation bias.

The second study (Buckreis, 1999) was a within-teacher comparison of a fourth-grade teacher, Meg, selected because she had experience with the fourth-grade curriculum but also because her understanding of one mathematical concept, multiplication, was much stronger than her knowledge of another concept, division. Buckreis evaluated Meg’s knowledge by asking Meg to list subtopics that fell within each of these two general curricular domains and to provide sketches or diagrams as needed to show how the topics related to one another. Meg’s knowledge of division was both faulty and incomplete on several topics, including the different meanings of division, the conceptual underpinnings of division procedures, and the idea of divisibility itself. The difference between her understanding of multiplication and of division made Meg an ideal candidate to study the influence of subject matter knowledge on teaching. In his observations, Buckreis saw that Meg did not provide students with a complete development of the full range of division situations. Moreover, at the conclusion of the observations, in a posttest, Buckreis found that students had significantly more success with multiplication problems than with division problems.

In the third study, Stein et al. (1990) studied a fifth-grade teacher, Mr. Gene, as he taught a 25-lesson unit on functions and graphing. To ascertain the teacher’s knowledge of this content, these authors relied on an interview and a card-sort task, which was also administered to a mathematics educator, so that the teachers’ knowledge could be defined relative to a presumed expert’s knowledge. The contrast allows them to see that Mr. Gene’s conception of functions was missing important ideas about functional relationships and that much of his thinking was relatively superficial. For instance, in the card sort, Mr. Gene sorted cards by their format (equations vs. graphs, for instance), whereas the mathematics educator sorted them according to the nature of the relationship that they portrayed, regardless of the format used to portray the relationship. However, the authors also suspected that the teacher’s understanding of this content had become somewhat distorted by a metaphor he adopted from his textbook. Mr. Gene’s textbook frequently presented problems based on a metaphor of a function machine that produced values of $Y$ for each value of $X$ it received. Students would be asked to predict outputs for
The problems also focused on point estimation more than on functional relationships as a whole. When plotting ordered pairs, Mr. Gene emphasized the benefit of graphs to check for errors in predicting individual values and gave less attention to the nature of the function itself. Although the authors thought that Mr. Gene’s representation of functions was influenced by the textbook metaphor, they also argued that his less sophisticated understanding of functions led to several missed teaching opportunities and led him to view the value of this unit in terms of strengthening arithmetic knowledge, rather than providing a groundwork for future content learning.

In a sense, these three studies all test a relatively commonsense hypothesis that “you can’t teach what you don’t know.” However, they also provide a more nuanced notion of what content knowledge actually looks like. In each case, the

<table>
<thead>
<tr>
<th>Citation</th>
<th>n</th>
<th>Teaching responsibility</th>
<th>Qualification of interest</th>
<th>Evidence of quality of teaching practice</th>
<th>Strategy for establishing link</th>
<th>Link established?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blasquez (1998)</td>
<td>3</td>
<td>Elementary mathematics</td>
<td>Structure of math knowledge, beliefs about pedagogy</td>
<td>Constructivist implementation of a prescribed lesson, student achievement</td>
<td>Between-teacher comparison</td>
<td>Visible influence compromised by possible confirmation bias, hindsight</td>
</tr>
<tr>
<td>Buckreis (1999)</td>
<td>1</td>
<td>Elementary mathematics</td>
<td>Subject matter structure</td>
<td>Classroom teaching, student learning of multiplication and division</td>
<td>Within-teacher comparison, multiplication vs. division</td>
<td>Visible influence</td>
</tr>
<tr>
<td>Stein, Baxter, &amp; Leinhardt (1990)</td>
<td>1</td>
<td>Elementary mathematics</td>
<td>Knowledge of functions and graphs</td>
<td>Representation of functions and graphs</td>
<td>Thick description</td>
<td>Visible influence compromised by possible hindsight, confirmation bias</td>
</tr>
<tr>
<td>Smith (1989)</td>
<td>1</td>
<td>Elementary science</td>
<td>Six domains of knowledge</td>
<td>Planning and teaching science lessons</td>
<td>Thick description, teacher attribution of learning</td>
<td>Visible influence</td>
</tr>
<tr>
<td>Magnusson, Borko, Krajcik, &amp; Layman (1992)</td>
<td>6</td>
<td>Middle school science</td>
<td>Conceptual knowledge, pedagogical content knowledge</td>
<td>Change in pupil knowledge</td>
<td>Between-teacher comparison</td>
<td>No visible influence</td>
</tr>
<tr>
<td>Cunningham (1995)</td>
<td>6</td>
<td>Secondary science</td>
<td>Sociological understanding of science</td>
<td>Pedagogy used in innovative curriculum unit</td>
<td>Between-teacher comparison</td>
<td>Visible influence compromised by possible confirmation bias</td>
</tr>
<tr>
<td>Gess-Newsome (1992)</td>
<td>5</td>
<td>Secondary science</td>
<td>Subject matter structure</td>
<td>Unit planning mostly, some other aspects of practice</td>
<td>Between-teacher comparison, teacher attribution of learning</td>
<td>Weak influence</td>
</tr>
</tbody>
</table>
missing knowledge was not discrete facts but, rather, seemed to be the structural relations among concepts. At the same time, two studies were susceptible to confirmation bias. Moreover, their references to what the teacher failed to say to their students are likely influenced by hindsight. The Buckreis study overcame both of these potential weaknesses by augmenting observation data with student test data.

All three studies could have been strengthened if they had more explicitly sought evidence regarding alternative explanations for what teachers failed to say. What if, for instance, Mr. Gene had been interviewed after a lesson in which he failed to say something, and he volunteered in the interview that in retrospect he realized he should have mentioned this issue? In this case, we would have to attribute the failure to speak as caused by something other than missing content knowledge.

**Elementary and middle school science.** Two studies examined the role of content knowledge in the teaching practices of elementary or middle school science teachers. The first study (Smith, 1989) described a single fifth-grade teacher as she taught a unit on light and shadow. Smith was specifically interested in the teachers’ knowledge of light and shadow but she also examined knowledge of classroom organization and management, children’s thinking, science teaching strategies, and curriculum along with her orientation to teaching science. Smith combined video-stimulated interviews with a teacher journal and close analyses of the lessons to show where and how each of these six kinds of knowledge influences the teacher’s practice. For instance, she described an occasion in which students were creating shadows of different objects by holding them in front of a light. One student said she could make the shadow of a piece of hair disappear by holding the hair close to the light. This student’s discovery contradicted the lesson’s messages that even the smallest objects block light and therefore create shadows. The teacher was able to spontaneously account for this apparent contradiction by reminding students that when objects are closer to the light, they make larger shadows, and when a hair is placed so near the light, its shadow is too large and diffuse to be visible. She then demonstrated that when the hair is pulled further away from the light, its shadow was visible. The context of this study—a professional development program and a researcher who was part of it—invites rival hypotheses about observer confirmation bias and possibly a teacher trying to accommodate the researcher’s values; however, the sheer volume of evidence presented, and the details provided, is persuasive enough to override these alternative explanations.

The second study (Magnusson, Borko, Krajcik, & Layman, 1992) examined the relationship between teacher knowledge and gains in student content knowledge in the context of a computer-based laboratory unit on heat energy and temperature. The study used a logic that is more akin to quantitative studies than to qualitative assessments of both teacher knowledge and student knowledge before and after the unit. The knowledge of six eighth-grade teachers and a sample of their students was assessed via structured tasks and interviews whose transcripts were coded for correct or incorrect statements and for misconceptions. The authors then present complex tables showing patterns of what teachers knew and what their students learned. For example, they show that when teachers offered incorrect knowledge or misconceptions on the structured tasks, their students learned less during the year. And when teachers lacked knowledge of the misconceptions their students might have, their students learned less. However, the single greatest predictor of student learning was not what teachers knew, according to this assessment, but rather the number and type of learning activities they said they provided for their students. Although it may be the case that these learning activities derived from teachers’ knowledge, they apparently did not derive from the knowledge that the researchers assessed. No clear path between knowledge and practice was established.

These two science studies offer an interesting contrast because they provide such different approaches to evidence and inference. Smith inundated us with examples of specific classroom events, providing a two-volume study of a single teacher, whereas Magnuson and colleagues provided only pre- and postinterviews.
with no observations. Smith’s study shows us where and how Meg’s knowledge influences her teaching practice, whereas Magnuson and colleagues show us a pattern of regularities that suggest student learning is more influenced by learning activities than by the teachers’ knowledge. Any number of interpretations can be offered for the Magnusson data—that the teachers’ knowledge really is not relevant, that teachers with greater or better knowledge were more likely to spend time teaching these topics, that teachers with greater insight into their students’ comprehension were more likely to teach the additional units, and so on. A more in-depth approach to qualitative work might have provided insights into these issues.

Secondary science. Three studies focused on secondary science content knowledge. In the first, Cunningham (1995) studied teachers’ understanding of the sociology of science—that is, how scientists interact with each other to create new knowledge. Thinking that a better understanding of these processes might lead to a more progressive approach to teaching, the author assessed teachers’ knowledge of the sociology of science, selected six science teachers who varied in their knowledge of this domain, and then observed them as they implemented two specific curricular units. All teachers had participated in a professional development program designed to give teachers more understanding of how social processes influence research production and to encourage teachers to use more group processes in their own science classes. Cunningham found that teachers with stronger sociological understanding used more group work in their classrooms and that they acknowledged a greater variety of sources of valid scientific information. Because all teachers had participated in the same program, accommodation bias should not be an issue. However, Cunningham’s findings might still be susceptible to confirmation bias, for she could have noticed more “good” practices when teachers had more knowledge and more “not-so-good” practices in other teachers’ classrooms. This hypothesis is made more likely by the fact that Cunningham stipulated at the outset that sociological understanding of science should influence teaching practice.

In the second study, Gess-Newsome (1992) examined “subject matter structure” (SMS), referring both to the central concepts and themes in biology and to how different specific topics relate to one another and to these central concepts. Focusing on five secondary biology teachers, Gess-Newsome deciphered the SMSes that were implied in their textbooks, the SMSes that were conveyed through teachers’ classroom presentations and discussions, and the SMSes that teachers revealed in a postinterview about the relationship among curricular units and the teachers’ rationale for a variety of other teaching decisions. Analyses of classroom lessons focused on explicaded relationships among curricular topics and on explicit themes. Ultimately, the author found that these observed SMSes were not strongly related to teachers’ own maps of the content. Only one teacher demonstrated a direct translation of his own understanding of the material into his instruction. The remaining four differed in their desire to provide such an organizing framework to their students. However, even when this was not their explicit goal, teachers with greater knowledge of relations among topics did spontaneously make integrative connections and provided real-life examples during their classroom lessons.

The third study of secondary science teacher knowledge focused on teachers’ understanding of the nature of scientific knowledge (NSK; Lederman & Zeidler, 1987). These authors recruited 18 experienced biology teachers and assessed their knowledge with a set of Likert-type scales asking teachers to agree or disagree with statements describing scientific knowledge. Items were then aggregated into subscales measuring teachers’ perceptions of the extent to which scientific knowledge is amoral, parsimonious, testable, tentative, creative, and parsimonious. Classroom observers had no knowledge of teachers’ scores on the NSK scales, thus protecting themselves from possible confirmation bias. Ultimately, these authors used quantitative procedures to test for relationships between teachers’ knowledge and their practices. They ranked teachers according to their scores regarding the nature of scientific knowledge and then looked at the patterns in the messages that different teachers conveyed to their students about these issues. From the classroom data,
they generated 44 categories of messages but found only 1 of them to be statistically significantly related to teachers’ NSK scores. Because statistical significance typically means there is 1 chance in 20 that the finding is because of chance alone, and because 44 tests were made by these authors, chance alone would lead us to expect at least two “significant” links in this set of data, thus suggesting that their measure of teacher knowledge was not related to any of their measures of classroom practice.

These three studies offer an interesting comparison in part because they focus on such different aspects of scientific knowledge and in part because they use such different approaches to ascertaining influence. Cunningham (1995) looks at the sociology of science, Gess-Newsome (1992) at how knowledge is organized, and Lederman and Zeidler (1987) at characteristics of scientific knowledge. The presence of such diverse approaches to defining and assessing content knowledge illustrates an advantage of studying teachers’ knowledge qualitatively, for qualitative approaches allow more nuanced notions of knowledge. Yet as a set, the findings do not provide a strong case. Cunningham found an effect of sociological understanding, but her finding is susceptible to confirmation bias, and the other two authors found little or no influence.

Summary of studies focusing on content knowledge. Two groups of studies focused on teachers’ content knowledge, one relying on stipulated differences in knowledge and the other involving direct assessments of knowledge. To reconcile these groups of studies with quantitative literature, I ask two questions. First, are the conclusions generally comparable? And second, do the qualitative studies live up to their promise and offer deeper or more nuanced understanding of the relationship, in this case, between teachers’ content knowledge and their classroom practices?

With respect to the first question, the findings are uneven across the set of studies. Three of 11 studies showed a clear and unambiguous influence whose validity was not threatened in an obvious way. Two others found an unambiguous lack of influence, and the rest either found weak influence or were compromised by threats to their validity. That these findings are as uneven as those of their quantitative brethren suggests that the problem of establishing this relationship may not be a methodological problem. However, one difference in study methodology is worth noting. The two studies that saw no visible influence were actually hybrid qualitative–quantitative studies. Both started with qualitative data but then coded and quantified it to a point where they could present tabular findings. It is possible that these efforts to standardize their data resulted in masking the very things the authors sought.

So how do these studies add to our understanding of the role of teachers’ content knowledge? They do offer a variety of insights and hypotheses for further investigation. For example, Stein et al. (1990) thought their teacher’s knowledge of functions and graphs had been influenced by the function machine analogy his textbook used as a pedagogical tool, thus hinting that practice may influence knowledge even as we seek evidence that knowledge is influencing practice. Other authors, such as Blasquez (1998) and Gess-Newsome (1992), found that teachers’ beliefs and goals for teaching content had greater influences on practice than did knowledge, a possibility that also raises questions about how knowledge, beliefs, and values may influence one another. Both of these insights also remind us that knowledge is not a fixed resource that teachers repeatedly draw from over time, as chalk is, but rather is constantly changing in response to context, experience, beliefs, and values. Perhaps the very question of how knowledge influences practice is ill conceived. It assumes knowledge is a relatively fixed entity that can be called on at will and overlooks the role of goals, beliefs, and spontaneous interpretations of events in determining what teachers do at any given moment.

Influence of Specific Teacher Education Courses

In addition to content knowledge, researchers and policy makers have been concerned about the value of that portion of teachers’ education that is typically labeled “teacher education.” Few quantitative studies address the merits of particular content within teacher education. Instead, they tend to focus on general categories of experiences, such as college majors or
advanced degrees. One recent study examines course-taking patterns for all teachers in Florida (Harris & Sass, 2006) and finds no evidence that preparation in teacher education itself contributes to student learning. This section reviews three studies of specific courses within teacher education programs and the section below reviews studies of whole programs or program segments. One advantage all of these studies have over quantitative studies is that they focus on people who have recently graduated from teacher education and are in that sense comparable. Quantitative studies often include whole populations of teachers, including those who graduated in the past year and those who graduated 30 years ago, so we are uncertain what exactly their education might have been like, and we are uncertain as to whether it is even reasonable to ask about influences of programs across such broad time horizons.

Three studies examined the influence of specific teacher education courses. One (Boedecker, 1997) examined the influence of science methods courses; the other two (Artiles et al., 1998; Causey, Thomas, & Armeto, 2000) examined the influence of courses that were designed to foster cultural sensitivity in their students. All three studies followed students from their teacher education programs into their first teaching jobs and looked for evidence of course teachings in either the teachers’ practices or their reasoning about their practices. The studies are summarized in Table 3.

Boedecker (1997) was interested in a secondary science methods course that emphasized the use of hands-on and inquiry-based pedagogies. Her study focused on three graduates of her own teacher education program, interviewing teachers directly about program influences and also observing them each four times. Boedecker saw some examples of constructivist approaches but more often saw lectures. In their interviews, teachers attributed their pedagogy to the science courses they had taken rather than their science methods courses. Boedecker believed that the science methods courses were more consistent with National Science Education Standards and speculated that teachers would be more able to implement these standards if their science teachers had used them.

The next two studies were interested in the effect of courses designed to alter cultural attitudes. In one study, Artiles and others (1998) followed two novice bilingual teachers who had

---

TABLE 3

<table>
<thead>
<tr>
<th>Citation</th>
<th>n</th>
<th>Teaching Responsibility</th>
<th>Qualification of Interest</th>
<th>Evidence of Quality of Teaching Practice</th>
<th>Strategy for Establishing Link</th>
<th>Link Established?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boedecker (1997)</td>
<td>3</td>
<td>Secondary science</td>
<td>Science methods courses</td>
<td>Observed practice and lesson plans</td>
<td>Follow-up, content match, teacher attribution</td>
<td>Weak influence despite possible observer bias, accommodation bias</td>
</tr>
<tr>
<td>Artiles, Barreto, Pena, &amp; McClafferty (1998)</td>
<td>2</td>
<td>Elementary bilingual</td>
<td>Course in multicultural education</td>
<td>Integration of knowledge and beliefs, translation into practice</td>
<td>Longitudinal</td>
<td>Weak influence despite possible accommodation bias, observer bias</td>
</tr>
<tr>
<td>Causey, Thomas, &amp; Armeto (2000)</td>
<td>2</td>
<td>Middle school social studies</td>
<td>Course in student diversity</td>
<td>Beliefs and values, pedagogy</td>
<td>Follow-up, content match</td>
<td>Weak influence despite possible accommodation bias, observer bias, confounded context</td>
</tr>
</tbody>
</table>
been students in a course that the authors themselves taught. The students were followed through 2 years of full-time practice. These authors conducted in-depth interviews and stimulated-recall interviews and also asked teachers to create concept maps on several occasions. The article presents great detail about the concept maps and how they changed over time. From these multiple sources of evidence, the authors concluded that some important transformations had occurred in their thinking over time but not in all domains. For example, teachers added knowledge details to their maps but often did not change superordinate beliefs, and their rationale for teaching decisions often did not match self-reported pedagogical beliefs. The authors concluded that the teacher education program gave them general ideas but not the tools they needed to cope with the demands of a culturally diverse classroom.

Finally, Causey et al. (2000) were interested in a middle school social studies methods course designed to foster a more culturally sensitive attitude. They identified a set of beliefs that entering students tended to hold and that they hoped to alter. In particular, they felt students did not appreciate their own privileged positions. In their study, they examined student reflections throughout the course and found that their beliefs did not change much. However, two students did appear to have genuinely changed, and so they followed these two students after they graduated. The follow-up consisted of one observation and a stimulated-recall interview each year. The authors found that one of the two students had reverted to her former, less culturally sensitive beliefs, but the other appeared to have sustained the changes she had made during the class. However, this teacher worked in a predominantly White, middle-class suburb—where her new beliefs may not have been challenged—whereas the teacher who reverted taught in a more diverse, lower-middle-class school. Ultimately, then, the course had no discernable effect on the practices of urban teachers.

Influence of Whole Teacher Education Programs

The value of the teacher education program as a whole is the most contentious part of debates about teachers’ qualifications. Although state requirements for certification are remarkably diverse, all require teachers to obtain a bachelor’s degree and to take some number of courses in teacher education. Surprisingly, very few quantitative studies have examined the value of teacher education programs per se and instead often limit their “teacher education” question to whether teachers are certified or whether they have obtained advanced degrees, even though requirements for certification and degrees include many features other than teacher education courses per se. Many literature reviews (e.g., R. Greenwald et al., 1996; Wayne & Youngs, 2003) also focus on credentials rather than on program characteristics. Only one (Zeichner & Conklin, 2005) reviews studies that contrast alternative approaches to teacher education itself. Zeichner and Conklin (2005) concluded, from their review, that the literature is too messy to yield any coherent conclusions. They were also disappointed that so much of the literature addressed structural differences (e.g., 4-year rather than 5-year programs) rather than substantive differences. This is an area, then, where qualitative research may shed light. All of the studies reviewed here examined the content actually taught to students and looked for evidence that this specific content influenced teachers after they took full-time teaching positions.

Just as it would be a mistake to assume that teachers’ education consisted solely of their specific teacher education program, so would it be a mistake to assume that any of these studies examines the impact of an entire educational program. Most study one aspect of the program, such as the science preparation program or the literacy preparation program. Nine studies are described here and are summarized in Table 4. These studies were quite various, focusing on elementary literacy, mathematics, and urban education and on secondary science, English, and physical education.

Elementary reading and literacy. Two studies followed teachers who had studied in elementary literacy programs. One study, the Beginning Teacher Study (Flint et al., 2001; Hoffman et al., 2005; Maloch et al., 2003) was largely quantitative (described in Hoffman et al., 2005) but included some qualitative reports as well.
<table>
<thead>
<tr>
<th>Citation</th>
<th>n</th>
<th>Teaching responsibility</th>
<th>Qualification of interest</th>
<th>Evidence of quality of teaching practice</th>
<th>Strategy for establishing link</th>
<th>Link established?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flint et al. (2001), Maloch et al. (2003)</td>
<td>42, 101</td>
<td>Elementary reading</td>
<td>Program in reading vs. elementary</td>
<td>Practices that reflect program concepts</td>
<td>Between-teacher comparison</td>
<td>Visible influence compromised by possible accommodation bias</td>
</tr>
<tr>
<td>Deal &amp; White (2005, 2006)</td>
<td>2</td>
<td>Elementary literacy</td>
<td>Constructivist program</td>
<td>Program influence on teacher decisions</td>
<td>Longitudinal, content match, teacher attribution</td>
<td>No visible influence but potential reverse confirmation bias</td>
</tr>
<tr>
<td>Cady (2006a, 2006b)</td>
<td>2</td>
<td>Elementary mathematics</td>
<td>Program in math education</td>
<td>Change in beliefs and in practice</td>
<td>Longitudinal</td>
<td>Influence compromised by potential professional development program</td>
</tr>
<tr>
<td>Schultz, Jones-Walker, &amp; Chikkatur (2008)</td>
<td>4</td>
<td>Elementary urban</td>
<td>Program in urban education</td>
<td>Specific pedagogy</td>
<td>Follow-up</td>
<td>Weak influence despite potential accommodation bias, confirmation bias</td>
</tr>
<tr>
<td>Adams &amp; Krockover (1997)</td>
<td>4</td>
<td>Secondary science</td>
<td>Program in science education</td>
<td>Evidence that teachers translated concepts into practice</td>
<td>Follow-up, content match, teacher attribution</td>
<td>Weak influence despite possible accommodation bias, misspecified influence</td>
</tr>
<tr>
<td>Grossman et al. (2000)</td>
<td>3</td>
<td>Secondary English, language arts</td>
<td>Program in English education</td>
<td>Evidence that teachers appropriated concepts</td>
<td>Longitudinal study, content match, other</td>
<td>Weak influence despite possible observer bias</td>
</tr>
<tr>
<td>Nolan (1995)</td>
<td>2</td>
<td>Physical education, any level</td>
<td>Physical education teacher education</td>
<td>Specific pedagogy</td>
<td>Follow-up, content match</td>
<td>No visible influence despite potential accommodation bias</td>
</tr>
<tr>
<td>Valli &amp; Agostinelli (1993)</td>
<td>1</td>
<td>Secondary mathematics</td>
<td>Teacher education vs. not</td>
<td>Classroom control, planning, teaching practices</td>
<td>Longitudinal study, before and after</td>
<td>Visible influence compromised by potential observer bias, confirmation bias, reactivity, confounded context</td>
</tr>
</tbody>
</table>
(described in Flint et al., 2001, and Maloch et al., 2003). For the qualitative component, researchers conducted structured telephone interviews with 1st-year teachers who had graduated from several programs identified as exemplary approaches to literacy teaching. They also interviewed teachers who had earned generic elementary credentials at the same institutions. The researchers wanted to learn if teachers’ self-reported practices differed as a function of the particular institution they attended or the type of program within the institution. The study included more teachers than most qualitative studies (42 in Flint et al., 2001; 101 in Maloch et al., 2003) but much less intense data collection. No face-to-face interviews were conducted nor any direct observation of practice. However the interviews were repeated three times during the course of the teachers’ 1st year of full-time teaching. Analyses of themes suggested that graduates from exemplary programs felt more confident in their ability to teach reading, were more likely to mention specific features of their programs that they valued, and were more likely to base instructional decisions on student needs rather than curricular mandates. Because these findings derive entirely from self-reported practices, they are highly susceptible to accommodation bias. All interviewees understood that they were being asked by representatives of their alma maters to evaluate the quality of their preparation programs.

The second study followed two teachers longitudinally and focused on the interaction between program influences, teachers’ beliefs, and teachers’ practices. Deal and White (2006) describe the evolution of teacher thought from preservice student teaching through the 1st year of professional practice, and Deal and White (2005) describe the 2nd year of professional practice. These researchers begin with the assumption that teachers’ a priori beliefs have a great deal of influence on practice and so focus their study on whether or how the teacher education program influenced teaching decisions relative to school context and personal beliefs. They interviewed teachers about influences on their decisions and then sorted these into main categories, for example, students, institutional rules or norms, program ideas, personal beliefs, and so on. During student teaching and 1st year of teaching, teachers made numerous references to their teacher education program as an influence, but these references virtually disappeared by the 2nd year of professional practice, and decisions seemed to be dominated by contextual influences. However, the program emphasized the importance of attending closely to student thinking, and in the authors’ analysis, students were categorized as part of the school context. Thus if teachers said their decisions were influenced by their students, the researchers categorized the influence as related to context rather than to the program. It is unfortunate that they did not present more specific examples of how teachers’ perceptions of their children influenced their decision making so that potential program influences might be more apparent. Ironically, this apparent lack of program influence may be a case of confirmation bias in the sense that the researchers expected to find stronger influences from context and prior beliefs than for their program.

Elementary mathematics. One study followed teachers over time from a program that emphasized cognitively guided instruction (Carpenter, Fennema, Peterson, Chiang, & Lof, 1989; Fennema, Franke, Carpenter, & Carey, 1993; Vacc & Bright, 1999). The authors followed 12 teachers from their student teaching through their 2nd year of professional practice using a series of successive “snapshots,” each of which included questionnaire data as well as interviews and observations. They then returned for a final snapshot as teachers began their 6th year of professional practice. These authors assumed that teachers’ beliefs would be powerful mediators of program messages and used formal measures to assess both teachers’ general epistemological beliefs and their beliefs about mathematics teaching and learning. The questionnaire data were analyzed quantitatively and reported in Cady, Meier, and Lubinski (2006a). Two teachers were then identified whose patterns of change were strikingly different from one another, and these teachers were the focus of a qualitative study reported in Cady et al. (2006b). The authors found that the teacher whose beliefs were least aligned with the program at the outset had actually adopted program-recommended practices while participating in the program but later reverted to traditional practice. The teacher
whose epistemic beliefs were more aligned with the program was slower to change but sustained her changes over time. However, this teacher also participated in a professional development program that shared the same goals as the preservice program, so it is difficult to make a case that the program had a clear influence on her practice.

**Elementary urban education.** One study (Schultz, Jones-Walker, & Chikkatur, 2008) followed teachers after they completed a preservice program that was explicitly designed to prepare them to teach in urban, high-poverty schools. The program advocated a “listening stance” to help their predominantly White, middle-class novice teachers learn to teach to students who were quite different from themselves. The study took place in a school with high teacher turnover and a rigid curriculum with a 6-week testing schedule, so teachers had few opportunities to tailor their curriculum to student interests. The authors focused on four program graduates teaching in this school, each of whom had 1 to 4 years of experience. They describe these teachers’ practices as deriving from a negotiation between the demands of the situation and their own values and beliefs. Teachers varied in how strongly they embraced the listening stance but generally found few ways to incorporate what they learned by listening into their curriculum. The authors attributed this difficulty in part to curricular constraints, but it is not apparent that the program offered any guidance on what to do with knowledge about students that was obtained by listening. The program may have been oriented more toward listening than toward incorporating what is learned from listening into curriculum and instructional decisions.

**Secondary science.** One study (Adams & Krockover, 1997) sought evidence of program impact by comparing the ideas of program graduates with ideas that were embodied in program rhetoric. The authors examined program documents and also interviewed faculty about key ideas, concepts, and beliefs that they wanted teachers to learn. They then visited four recent graduates (with 1 to 3 years of experience) who had also been interviewed during their student teaching and were known to hold different perceptions of the relative value of their teacher education courses. Interview questions targeted the source of teachers’ knowledge, and the authors used responses to trace teachers’ ideas back to specific courses. They found that some aspects of the program did translate into practice and did become part of teachers’ knowledge. However, the program’s influence, especially its constructivist ideas, was often neutralized by competing influences from their own prior beliefs, their classroom experiences, and their school contexts.

**Secondary English.** Two studies looked at the effect of secondary English programs. The first study was presented in a dissertation (Grossman, 1988b), a book (Grossman, 1990), and three journal articles (Grossman, 1988a, 1989; Grossman & Richert, 1988). The study examined six secondary English teachers, three of whom had not studied teacher education and three of whom had. The author did not follow teachers longitudinally through the program but did try to match teachers on their general educational background. The study focused on teachers’ pedagogical content knowledge and provided a sophisticated theoretical rationale for its design and focus. To ascertain teachers’ pedagogical content knowledge, the author used a series of structured interviews about literature and teaching literature, the teachers’ perceptions of the courses they took as part of their preservice program, how they planned their own courses, and their rational for a particular unit of instruction. She also observed teachers repeatedly during a single unit of instruction. The author found that the three graduates of teacher education held similar ideas about the purposes for teaching English and that they frequently attributed their ideas to their subject-specific course work. They also chose a wider variety of literature than the other three teachers. The teachers without professional education found it more difficult to anticipate what students would find difficult in English and that they frequently attributed their ideas to their subject-specific course work. Though the study provides substantial detail and a strong theoretical rationale, it remains susceptible to both accommodation bias and confirmation bias. In addition, the author documented program messages after these students had graduated, so there is a possibility that the program the author saw was not the same as the program teachers saw.
The second study of elementary and secondary English teachers (Grossman et al., 2000) involves a longitudinal study of 3 novice teachers followed from their student-teaching year through their first 2 years as professional teachers. The authors set out specifically to address the problem of why teacher education programs lack a stronger influence on practice. They studied 3 teachers selected from a larger sample of 10 but do not say how either the 10 or the 3 teachers were selected. Data collection consisted of five observations and interviews per year. The authors found that teachers adopted instructional concepts, such as scaffolding, a process orientation to writing, and the value of ownership in writing, from their teacher education programs but that they had to go elsewhere to find the practical tools they needed to actually teach. They depended heavily on district curriculum materials even when these were antithetical to the program’s concepts yet did not abandon program concepts. From these observations, the authors conclude that teacher education can play an important role in helping teachers, for instance, by providing teachers with a vision of writing instruction, but that teachers would benefit more if programs also provided specific instantiations of how these visions could be realized.

Physical education. Nolan (1995) was interested in whether graduates of her physical education teacher education program continued to use the methods they learned in the program after they had been practicing for 6 to 8 years. Nolan asked two former students to videotape themselves teaching and then mail back the videotapes. The author still had videotaped data of these teachers from when they were student teaching and was able to do a direct comparison of “before and after” teaching. She also conducted telephone interviews with each teacher. She found that both teachers still used, albeit selectively, practices they had learned in the program but that in general, they engaged in less instruction and more management than they had during their student teaching and that both were more concerned with student compliance and enjoyment than with student learning. Both teachers indicated in interviews that student learning was not a priority for administrators, the public, students, or even other faculty. Instead, the general press was toward behavior management. Nolan speculated that these contextual influences likely motivated the changes in practice.

Secondary math. One study (Valli & Agostinelli, 2000) examined a single secondary math teacher who was also the second author of the research report. This teacher had taught before completing his baccalaureate degree and without having taken any teacher education courses. He then returned to college, completed a teacher education program, graduated, and returned to teaching. The study compares the quality of his teaching before and after he enrolled in the teacher education program and concludes that he had substantially improved his practice. However this finding is compromised by numerous alternative explanations. For instance, the teacher taught in a different context (a public high school) after participating in the program than he had taught in before (a Catholic high school), and the observations of his initial teaching were made by a student teacher, not a researcher, and reflect the impressions of an inexperienced teaching candidate. The study is susceptible to accommodation bias, confirmation bias, confounding of observers over time, and confounding of teaching context. One could argue that because the teacher is a coauthor of the article, the possibilities might be mitigated, but there is no reason to believe that participation as a researcher would necessarily reduce personal investment in the findings. The sum of these problems raises doubts about the validity of the findings.

Summary of studies of teacher education. These studies of the influence of teacher education courses and programs are quite various, not just in the domains they examine but also in the outcomes they seek and in their research methods. Still, as a group, they differ from those that examine content knowledge in that none of the studies of teacher education found a clear, visible influence on teaching practice. One possible reason for this difference is that the types of outcomes they seek are more difficult to achieve. Most of these programs want to alter the things teachers do, such as using an inquiry approach to teaching science or adopting a listening stance in an urban school. These ideas are not
very clearly defined, and even though teacher educators describe them, their students may be unable to envision them or to appreciate the specific tasks needed to pull them off. Moreover, these approaches to teaching require a lot of spontaneous mid-lesson adjustments, things novices are not good at. Kennedy (1999) refers to this problem as the problem of enactment. So just as findings from quantitative studies are quite various and lead to different conclusions, as the quoted passages at the beginning of this article suggest, so do these qualitative studies tend to yield ambiguous conclusions.

But the qualitative studies do allow us to see more clearly why program influences yield the patterns that they do. Several themes appear across these studies. One is that teachers are very strongly influenced by the specific circumstances in which they find themselves—the students they serve, the curriculum and other materials at their disposal, organizational constraints and norms in their buildings. Another is that program influences are mediated by teachers’ prior beliefs and sometimes are entirely revised by these beliefs. Indeed, these two themes are so pervasive that the more recent publications reviewed here begin with the assumption that program influences will be overwhelmed by these other influences. The third important theme is that program content tends to lack concrete examples. Several authors found that teachers claimed to embrace program concepts but that they were unable to translate those ideas into practice.

These themes are not new to most teacher educators. Studies of student teaching conducted in the 1970s pointed to the strong influence of local contexts as counteracting the influence of teacher education programs (see, e.g., Hoy & Rees, 1977; Ryan et al., 1979; Zeichner, 1980). I eliminated studies of student teaching from this review, but the similarities between the literature reviewed here and the literature on student teaching suggest that the teacher education community has been aware for several decades that its programs were not powerful enough to alter school norms.

**Discussion**

The current spate of arguments about the knowledge needed for teaching, and about the value of teacher education for teaching, began in earnest after the National Commission on Teaching and America’s Future (1996) released its report, “Teaching and America’s Future.” That report spawned a vigorous debate about the importance of teacher education and certification (Ballou & Podgursky, 1998, 2000a, 2000b; Darling-Hammond, 2000) that continues today. The studies reviewed here are accompanied by hundreds of quantitative studies that have also sought associations between teachers’ knowledge or educational backgrounds and the quality of their practice. Each set of studies is susceptible to its own alternative explanations of findings, so each literature is difficult to review and summarize.

Qualitative studies offer several advantages for looking at these relationships. Qualitative researchers tend to have more detailed and nuanced assessments of teachers’ knowledge and to know more about what teachers actually learned in their educational programs and when they completed those programs. Quantitative researchers often have only rough indicators of knowledge: tallies of courses taken or test scores but no knowledge of the content of those courses or tests. Moreover, quantitative samples often include teachers who completed their educational programs at wildly different times in the past. That the two sets of studies yield messy and ambiguous conclusions attests to the difficulty of sorting out causal influences in a dynamic world.

It should be noted, too, that the value of qualitative studies for questions of causal influence has received relatively less attention, and as a result, less attention has been given to explicit design decisions. I have tried to facilitate progress along these lines by outlining some of the prominent alternative explanations that researchers need to anticipate and control. Probably the most disappointing aspect of these research reports is their lack of detail about sampling rationale. Yet when samples were justified, their rationales contributed to our grasp of the findings. For instance, when Causey et al. (2000) say that the two teachers they followed were the only two who demonstrated change during their participation in class, we understand that their transition into professional practice could represent the best possible program outcome, not a random program outcome. When Cady et al.
Kennedy (2006b) tell us they followed teachers whose initial beliefs were differentially aligned with program goals, we understand that we will be learning how the program’s messages become differentially translated into practice by these two teachers. But when Grossman and others (2000) say they studied 3 teachers from a sample of 10, without saying how they selected either the 10 or the 3, we do not know if the sample represents the best case of learning from teacher education or if it represents some particular kind of learning from teacher education.

One also wonders how much more might have been learned from these studies if qualitative researchers were more self-conscious about controlling for confounding variables, documenting contexts and mitigating circumstances, and taking a more skeptical stance toward their own interpretations and evidence. Fifteen of the 23 studies reviewed here were compromised by design flaws that raised questions about their findings, thus limiting their potential to help us understand how teachers’ knowledge and credentials help them teach. Notice, too, that the limitations I have outlined are not limitations inherent in qualitative research. They do not suggest that qualitative research is inherently weaker than quantitative research. But they do demonstrate that when qualitative researchers establish truth-seeking goals, they need to pay more attention to their own methods.

More important than their limitations is the messages about the role of knowledge, credentials, and teacher education that these studies offer. Because they can look at practice as it unfolds, they offer observations that quantitative studies can rarely provide. One important message that these studies provide has to do with the tremendous power of classroom and school contexts over practice. Schools structure teaching practice through schedules, textbooks, materials, and rules. Students also influence teaching practices. They are the primary audience from whom teachers can get feedback about their work. The problem of situational press is not unique to teaching, of course. We all forget about admonitions to eat less sugar and fat when we are hungry and surrounded by vending machines full of potato chips and candy bars.

Another important message here is that studies seeking evidence of a role for content knowledge were more successful than those seeking evidence of a role for teacher education programs. This is somewhat of a surprise, for quantitative studies have been more equivocal about the role of content knowledge. That studies of teacher education programs were so rarely able to reveal any clear and visible influences is particularly surprising, because most of this literature was generated by people who were themselves teacher educators who were deeply knowledgeable of their courses and programs and thus, presumably, would be more able to see influences if they were there.

The third message offered by these studies is that teacher education courses provided only general ideas or concepts, without the kind of detailed guidance teachers need to translate these concepts into specific strategies. The studies introduce us to a teacher with extensive field experience in earth science who is unable to translate that experience into engaging lessons, despite having completed a teacher education program, and to teachers who learned to listen to their students but did not know how to incorporate what they heard into their lessons. Teaching entails translating ideas into events, and many of the teachers described here were unable to do that. We could speculate, as some of these authors did, that teacher education could be more influential if it provided examples of how to translate its ideas into specific events. This hypothesis reinforces arguments for the use of such tools as cases and hypermedia to examine specific lessons, student work, or school contexts (for a review of this work, see Grossman, 2005).

Finally, several studies suggest that there is a two-way interaction between knowledge and practice. Knowledge did not remain fixed in teachers’ minds after they left college but instead continued to evolve as it interacted with both beliefs and teaching experience itself. Most quantitative work is based on an assumption that the knowledge measured by licensure tests, or instilled by curriculum requirements, remains unchanged over time and that it is just as relevant to a 30-year veteran as it is to a new teacher. But if knowledge continues to evolve over time, we would not expect to see a relationship between credentials or college course work in a random sample of teachers who graduated at different
times in the past. So knowledge and program messages, both intended to be used to control events in the classroom, are themselves influenced by these same events.

These studies might also provide an important stimulus for teacher educators. They demonstrate many reasons why program ideas may never find their way into classroom practices and in so doing may stimulate a new generation of hypotheses about how programs can have a greater influence, and a new generation of programs that are more grounded in school life.

References

References marked with an asterisk were included in the meta-analysis.


*Carlsen, W. S. (1997). Never ask a question if you don’t know the answer: The tension in teaching between modeling scientific argument and maintaining law and order. Journal of Classroom Interaction, 32(2), 14–23.


**Authors**

MARY M. KENNEDY is a professor of education at Michigan State University, College of Education, 116 Erickson Hall, East Lansing, MI 48824; e-mail: mkenndey@msu.edu. Her scholarship focuses on defining teacher quality and on identifying the things that most influence teacher quality. She has examined the influences of teacher education, research knowledge, attitudes and beliefs, credentials, and school context.

Manuscript received November 21, 2007
Final revision received August 18, 2008
Accepted September 3, 2008