Teaching is the sort of work that both inspires and mystifies, and many people have tried to articulate the qualities they believe make someone a good teacher. Some say a good teacher is bright, others that she is caring, others something else. These speculations take on a more practical quality when the conversation turns to how to prepare good teachers, for the preparation question can lead to detailed curriculum and program specifications. So literature on the question of how to produce good teachers includes philosophical inquiries, program designs, and empirical tests of hypotheses.

Arguments about how to prepare teachers have become especially shrill in recent years, as observers become increasingly concerned about the quality of the education system as a whole. Debates have also been stimulated by a new body of work called “value-added” analyses, which examine variations among teachers’ classroom effectiveness. From these analyses, we know that teachers vary considerably in the amount their students learn, so much so that if a student had two or three consecutive weak teachers, his overall academic achievement would be seriously compromised (Aaronson et al., 2003; Rokoff, 2003; Sanders & Horn, 1998). These studies demonstrate that teachers differ substantially in their effectiveness, and raise to prominence the question of how to better prepare teachers.

The aim of this paper is to examine empirical evidence regarding the merits of the most prominent hypotheses about educational backgrounds that will improve teacher effectiveness. One hypothesis, the one that dominates most state regulations, is that teachers need specialized knowledge about issues directly pertinent to teaching—things like classroom management, techniques for teaching, the role of school in society and other educational issues. We call this hypothesis the Pedagogical Knowledge hypothesis. Virtually every state subscribes to this hypothesis by requiring prospective teachers to take courses from departments of teacher education whose mission is to prepare people specifically for teaching careers. But even though this hypothesis is widely represented in state regulations, it is not without its detractors. In particular, two other hypotheses are offered, either as alternatives or as supplements. One argues that teachers need Content Knowledge more than pedagogical knowledge. Proponents of this hypothesis frequently note that teachers cannot teach content if they do not know it. While it is possible in principle for teachers to obtain pedagogical knowledge as well as content knowledge, advocates for content knowledge frequently pit the two against one another under the assumption that courses providing pedagogical knowledge take up too much space in the college curriculum and hence remove space for courses providing content knowledge that would ultimately be more beneficial to prospective teachers. There is, among those outside the teacher education community, a general skepticism about the merit and value of courses taken in teacher education programs per se (Conant, 1963; Damerell, 1985; Hess, 2001; Kramer, 1991; Labaree, 2004; Lagemann, 1999).
In the past two decades, a third hypothesis has been put forward that suggests that teachers need a blend of pedagogical and content knowledge, something called Pedagogical Content Knowledge. This knowledge consists of such things as how students understand, or misunderstand, particular substantive ideas, how to present particular substantive ideas in a way that makes them more accessible to different types of students, or how to use particular resources in lessons about particular content. This third hypothesis, then, suggests that there is a relationship between pedagogy and content that teachers need to understand.

All of these hypotheses focus on the college curriculum, suggesting that there are particular domains of knowledge that can make a difference, and that teachers should take courses in these domains. There is another hypothesis that challenges the validity of all of these. This fourth hypothesis argues that the best teachers are Bright, well-educated people who are smart enough and thoughtful enough to figure out the nuances of teaching in the process of doing it. For people subscribing to this hypothesis, the route to improving the quality of teaching lies in recruitment, not in specific courses that will prepare people for this work.

Advocates of the first hypothesis tend to acknowledge all of the rest as well. That is, they rarely argue against content knowledge, against pedagogical content knowledge, or against the value of having bright, well-educated people teaching in the nation’s classrooms (see, e.g. National Commission on Teaching and America’s Future, 1996). However, advocates for content knowledge and for bright, well-educated people often do argue against the pedagogical knowledge and against pedagogical content knowledge hypotheses. So the debate really focuses mainly on the merits of allocating college curriculum space to courses specifically about teaching and learning.

We take the view that all of these hypotheses need to be examined. Therefore, we seek evidence about all of them, rather than confining ourselves to the most controversial one. For the first three hypotheses, we examine studies of teachers’ college course-taking histories. We assume that courses in education represent pedagogical knowledge, courses in mathematics represent content knowledge, and courses in mathematics education represent pedagogical content knowledge. Because our fourth hypothesis deals more with recruitment than with college curriculum per se, we test this not by looking at studies of college courses, but instead at studies that examine the selectivity of the teachers’ alma maters.

In this investigation, we confine our inquiry to the content area of mathematics, and take student achievement in mathematics as our indicator of teachers’ effectiveness. This limitation is necessary in part to ensure that all curricular options are tested against a common outcome. We test hypotheses only about teachers’ educational background, not about their tested knowledge.

The studies reviewed here come from a larger collection of literature addressing the role of a wide range of teacher qualifications to the quality of teaching. Below we describe our search procedures for gathering this literature and discuss some of the methodological problems in this literature. We then review findings for college curricula and for recruitment.

**LITERATURE SEARCH PROCEDURES**

Literature for this paper was drawn from a larger literature data base gathered as part of the Teacher Qualifications and the Quality of Teaching (TQQT) study. The TQQT data base includes studies that examine the relationship between at least one teacher
qualification and at least one indicator of the quality of teaching. Our original search criteria defined qualifications to include such aspects of teachers’ backgrounds as their college curricula, test scores, credentials, grade point averages and degrees. Indicators of teaching quality included such things as direct observation of classroom practice, student achievement, and principal ratings, as long as these indicators were obtained after the teacher had a full time teaching position.3

Excluded from this compilation were studies of student teachers, preschool teachers, and teachers of college students and adults. Also excluded were studies published prior to 1960 and studies conducted outside the United States, on the grounds that these contexts may be too different from contemporary United States to be applicable. For our tests of the four hypotheses about educational backgrounds, we also eliminate studies published before 1980, on the grounds that curricula change over time and earlier studies may no longer be applicable.

Literature was obtained by searching the Education Resource Information Center (ERIC), PsycInfo, Dissertation Abstracts International, and EconLit. Search terms included those commonly used to define either qualifications of teaching quality, such as assessment, certification, teacher education, teacher effectiveness and so forth. In addition, we searched the bibliographies of these articles as well as those of literature reviews and policy analyses in this area and searched recent issues of entire journals whose domain encompassed this area. Studies were screened to ensure that they included at least one qualification and at least one indicator of quality, and that their analysis explicitly linked the two. Links could be established with group comparisons (e.g. a group of traditionally certified teachers versus a group of alternatively certified teachers); correlations, various multivariate strategies, or through qualitative approaches. As of this writing (Summer 2006) the data base included about 480 studies. More details about the study and the data base can be found at http://www.msu.edu/~mkennedy/TQQT.

To test our four hypotheses about teachers’ educational backgrounds, we searched this literature for two types of studies. First, to test hypotheses about the college courses, we sought studies that tried to assess the relationship between the courses teachers took in college and the teachers’ current effectiveness. We sought tallies of courses taken in education, mathematics, or mathematics education. We also sought tallies of whether teachers had majored or minored in either of these subjects or held advanced degrees in them. The bright, well-educated person hypothesis cannot readily be tested by examining college curricula, but we did find two types of studies that are relevant to that hypothesis: those that measure of the status, or selectivity, of the institution which teachers had attended for their college education, and those that looked at the effectiveness of teachers who were recruited by Teach for America. Because the TFA recruits from prestigious colleges and universities (Raymond & Fletcher, 2002b), we consider these studies to be tests of the bright, well-educated person hypothesis.

Methodological issues

Research on the relative value of different kinds of teacher preparation has been impeded by some very difficult methodological problems. One problem is that many of the credentials of interest don’t really vary much. For example, the value of a bachelor’s degree would be difficult to assess in the United States because some 99 percent of teachers already have a bachelor’s degree (National Center for Education Statistics, 2003). A test of the value of this degree would require an extensive search to find teachers who lacked the degree, to serve as a comparison group for those who have it. In many important respects, the teaching population in the United States is quite homogenous, a
phenomenon that is hard to reconcile with the large differences in effectiveness that we see. For example, we know that about 75 percent of them are women and 84 percent are white (National Center for Education Statistics, 2003). We also know that 99 percent of teachers have a bachelor’s degree and that 90 percent of teachers are certified.

On the other hand, bachelor’s degrees and teaching certificates can have quite different meanings across educational institutions and across states. The United States has hundreds of institutions offering college degrees and they vary in size, quality, religious affiliation, research orientation and so forth. In addition, states have a plethora of certifications on their books, often exceeding a hundred. And they differ in the curricula they require for any given certificate (Ballou & Podgursky, 1999; Council of Chief State School Officers, 1988; Rotherman & Mead, 2003). We avoid ambiguous measures such as certification and degree level and focus instead on the number of course teachers take in particular subjects and on the selectivity of their alma mater. Of course these also can have different meanings in different places but we think their meaning is more clear than certification is. This strategy helps increase the variability in measured educational background, and at the same time improves the clarity of their meaning.

Another problem affecting all studies of qualifications is that teachers decide for themselves which qualifications they will obtain. Researchers do not randomly assign teachers to colleges or universities, nor to their college majors. Thus measurements such as the number of courses taken on topic X or Y reflect not only the knowledge teachers gained from these courses but also their initial interest in those topics. We can never separate teachers’ initial interest in a subject from the knowledge they have acquired about it. Thus, if we find that courses in a certain topic seem to lead to better-quality teaching, we cannot infer that if states require teachers to take those courses, that the same results will obtain. Only after the courses are required can we see their effect on teachers who have no interest in them. This problem plagues the present examination as much as every other examination of teachers’ qualifications. Prospective teachers take some college courses because they are required to, but they take others because they want to, and it is likely that these two reasons bear on the degree of benefit that teachers derive from their courses.

A similar self-selection phenomenon occurs when teachers seek employment, for their assignment to schools is also not at all random. In fact, there is now considerable evidence regarding the way teachers and schools become matched. We know, for instance, that most young people who choose teaching as a career are white women who come from rural and suburban communities. We also know that, upon graduation, they seek positions in schools that are similar to those they attended when they were students themselves (Boyd et al., 2003). They engage in a process of differential migration, such that college graduates who grew up in the suburbs seek teaching positions in the suburbs, and those who grew up in small towns seek teaching positions in small towns. We also know that school systems give preferences to their own graduates (Strauss, 1999), so that both schools and teachers are seeking to match each others’ cultural and demographic fundamentals. Finally, we know that urban schools, and other schools serving lower income and non-white populations, have greater difficulty filling their vacancies, have higher turnover in their teaching staffs, and are more likely to employ teachers with fewer qualifications (Boyd et al., 2002; Lankford et al., 2002; Wykoff, 2001). As a result they tend to have more novices than other schools and to have less-qualified teachers as well. These processes, taken together, suggest that the positions teachers eventually take are likely to be affinity assignments, such that their social backgrounds and their qualifications match the social backgrounds and qualifications of their students.

Now imagine how these processes influence research. We want to see if teachers with different types of qualifications have different influences on their students. But teachers
with different qualifications have already been matched to students with different qualifications, through these natural processes of differential migration and affinity assignments. One result of these processes is that the schools whose students need the most educational help—that is, those who are the least qualified as students—are allocated the least qualified teachers, while those who are the most advantaged are allocated teachers with more qualifications. If we examine a simple correlation between, say, the prestige of the teachers’ alma mater and their students’ achievement test scores, we would likely see a relationship even if the teachers have had no impact at all on their students, simply because they have been matched to their students through these processes of differential migration and affinity assignments. The migration and assignment processes make it difficult for researchers to tell whether teachers created their students’ achievement levels or whether, instead, students with different achievement levels have attracted different kinds of teachers.

We address this problem of non-random assignment by limiting our review to studies that use relatively sophisticated statistical methods. We seek studies in which researchers try to separate out these various influences statistically, by measuring all the things they expect to influence teachers’ effectiveness, and include them in a statistical model. These more sophisticated models include not only teachers’ qualifications, but also measures of students’ socio-economic status, race or ethnicity, perhaps measures of school size, finances, or demographic makeup, and perhaps other characteristics of the teachers themselves, such as their beliefs or values. In these studies, the goal is to find a relationship between teachers’ qualifications and the quality of their teaching practice after measuring and accounting for other relevant influences. They cannot guarantee that they are documenting causal influences, of course, but they are more able to measure and take into account competing explanations for their data. We also strengthen our inference by examining multiple studies, each of which uses different statistical models.

Finally all studies of teachers’ qualifications are complicated by numerous influences that intervene between teachers’ prior college experiences and their current teaching practices. Any given sample of teachers will include some relatively new teachers, some with a few years of experience and some with many years of experience. Many things may have influenced their teaching since they obtained their original college degrees and certifications. They may have worked in different schools, taken different kinds of professional development courses, confronted parents with different expectations for their children. Any of these experiences may have influenced their practices and their effectiveness. Yet our research is ignorant of these influences and seeks evidence of only one influence, their college education, and that may have occurred years ago. Indeed, time itself is an intervening influence, for teachers who obtained their certificates 20 years ago may have experienced very different educational programs than teachers who obtained their certificates five years ago. And studies that were published 20 years ago may have included teachers whose preparation was quite different from that of teachers prepared today (Metzger et al., 2004).

We address these problems in part by limiting ourselves to stronger research studies, studies that are more likely to have addressed these issues. We include only those studies that meet these important criteria.

1 The studies must use relatively sophisticated models, models that allow the researcher to take account of numerous other possible influences. These studies typically rely on either multiple regression or hierarchical linear models, both of which allow more opportunity than do simple correlations or group comparisons to measure additional influences and take them into account.

2 The studies must include a pretest as one of their factors. Measures of student
achievement are like snapshots taken at a specific time. They measure student knowledge at a particular time, but that knowledge reflects the sum of everything students have learned throughout their entire lives, from parents, other teachers, other adults and from each other. Moreover, they can present a relationship that reflects the effect of affinity assignments more than the effects of teachers’ influence on student learning. We therefore prefer studies that incorporate a pretest score as one of the factors in their model, or as part of a gain score. The inclusion of some measure of prior achievement enables the researcher to separate the knowledge students gained prior to studying under the teacher who is the focus of the study.

3 The studies must focus on individual students or teachers rather than institutional collections of teachers. Some researchers rely on school averages or district averages of teachers and students, rather than using data from individual teachers. Aggregated data are often used when researchers borrow data bases that belong to school districts or states. Such data bases may not include individual-level information, or may restrict access to it to protect teachers’ privacy. These aggregated data conceal all of the variations among teachers within each school or district and focus on variations among schools that may be due to many things other than instructional influences, things researchers have no knowledge of (for more on this issue, see Murnane, 1981; and Hanushek et al., 1996). As a result, they may lead to erroneous estimates of the relationship between teachers’ qualifications and the quality of teaching practices. We therefore limit our attention to studies that focus on individual teachers and students.

One final note is needed about the findings we present here. A synthesis of this sort requires us not only to define our criteria for including studies, but also to define criteria for selecting specific findings from each study. Researchers frequently present multiple statistical models, each with unique features. They sometimes provide findings for multiple subgroups (e.g. high-achieving and low-achieving students) or for multiple achievement sub-tests. To avoid over-representation of any one study, we selected a single model from each grade or sample examined. Our decision rules were to first seek models that met our criteria of including a pretest and focusing in individual teachers rather than aggregations. If an author presented one model using a gain score and another using the pretest as a covariate, we chose the covariate model. We selected fixed effects over random effects. We selected total samples rather than subgroups and total test scores rather than sub-test scores. We avoided models that included interaction terms. If different models from the same sample are presented in different publications, we selected the most recent version. Otherwise we selected the model that included the largest number of curriculum measures or, absent differences in this, the one that was most complete with respect to other variables. We preferred one-year gains to two-year or partial-year gains.

Final sample

The studies we examine here represent what we consider to be the best evidence of the relationship between teachers’ qualifications and their students’ mathematics achievement. By limiting our review to these studies we gain more confidence in our findings. Our criteria have been recognized by numerous other researchers as indicators of better research designs (see, e.g. Goldhaber & Anthony, 2003; Hanushek, 1971; 1996; Murnane, 1981). In this respect, we follow Slavin’s (1984, 1986) suggestion that reviewers should focus on best evidence, and we agree with him that what constitutes best evidence depends on the research questions and contexts. Table 61.1 lists the studies that meet our criteria for best evidence.
Table 61.1  Studies included in this synthesis

<table>
<thead>
<tr>
<th>Citation</th>
<th>Educational background of interest</th>
<th>Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hypothesis</td>
<td>Indicator</td>
</tr>
<tr>
<td>Aaronson, Barrow, &amp; Sander (2003)</td>
<td>CK PK</td>
<td>Major in: Math Education</td>
</tr>
<tr>
<td></td>
<td>BWE</td>
<td>Level of: Institutional status</td>
</tr>
<tr>
<td>Betts, Zao, &amp; Rice (2003)</td>
<td>CK BA in: Math</td>
<td>Elem</td>
</tr>
<tr>
<td></td>
<td>PK</td>
<td>Middle</td>
</tr>
<tr>
<td></td>
<td>PK</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Minor in: Math</td>
<td>Elem</td>
</tr>
<tr>
<td></td>
<td>PK</td>
<td>Middle</td>
</tr>
<tr>
<td></td>
<td>PK</td>
<td>High</td>
</tr>
<tr>
<td>Brewer &amp; Goldhaber (1996); Goldhaber &amp; Brewer, (1996a, 1997a; 1999; 2000)</td>
<td>CK PK</td>
<td>Major in: Math Education</td>
</tr>
<tr>
<td></td>
<td>PK</td>
<td>Middle</td>
</tr>
<tr>
<td></td>
<td>PK</td>
<td>High</td>
</tr>
<tr>
<td>Chiang (1996); Rowan, Chiang, Miller (1997)</td>
<td>CK Double degree in: Math</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>PK</td>
<td>Any degree in: Math</td>
</tr>
<tr>
<td>Darling-Hammond, Holtzman, Gatlin, &amp; Hellig (2005a; 2005b)</td>
<td>BWE Program entry: TFA</td>
<td>4–5</td>
</tr>
<tr>
<td>Decker, Mayer, &amp; Glazerman (2004)</td>
<td>BWE Program entry: TFA</td>
<td>1–5</td>
</tr>
<tr>
<td>Eberts &amp; Stone (1984)</td>
<td>PCK N of courses in: Math Ed</td>
<td>4</td>
</tr>
<tr>
<td>Fagnano (1988)</td>
<td>CK PK</td>
<td>N of courses in: Math Math Ed Education</td>
</tr>
<tr>
<td></td>
<td>PK</td>
<td>Math Ed</td>
</tr>
<tr>
<td></td>
<td>PK</td>
<td>6–8</td>
</tr>
<tr>
<td></td>
<td>CK</td>
<td>9–10</td>
</tr>
<tr>
<td></td>
<td>CK 4–5</td>
<td>74103</td>
</tr>
<tr>
<td></td>
<td>6–8</td>
<td>116673</td>
</tr>
<tr>
<td></td>
<td>9–10</td>
<td>83516</td>
</tr>
<tr>
<td></td>
<td>PK 4–5</td>
<td>74103</td>
</tr>
<tr>
<td></td>
<td>6–8</td>
<td>116673</td>
</tr>
<tr>
<td></td>
<td>9–10</td>
<td>83516</td>
</tr>
</tbody>
</table>
(Continued Overleaf)
We have devised a novel yardstick for defining influences on student test scores. We want to see whether teachers who took more courses have students with higher achievement gains than other students. But the apparent influence of a college course will look different when student achievement is measured on a 50-point scale than when it is measured with a 10-point scale. The conventional solution to this problem is to divide all effects by the group’s standard deviation, thus creating a scale that puts all effects on a common

<table>
<thead>
<tr>
<th>Citation</th>
<th>Educational background of interest</th>
<th>Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hypothesis</td>
<td>Indicator</td>
</tr>
<tr>
<td>PCK</td>
<td>Pedagogical content</td>
<td>4–5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6–8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9–10</td>
</tr>
<tr>
<td>PK</td>
<td>Major in: Education</td>
<td>4–5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6–8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9–10</td>
</tr>
<tr>
<td>PCK</td>
<td>Math Ed</td>
<td>4–5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6–8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9–10</td>
</tr>
<tr>
<td>CK</td>
<td>Math</td>
<td>4–5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6–8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9–10</td>
</tr>
<tr>
<td>Hill, Rowan, &amp; Ball (2005)</td>
<td>CK+ PCK</td>
<td>N of courses in: Math and/or Math Ed</td>
</tr>
<tr>
<td>Monk (1994); Monk &amp; King (1994)</td>
<td>CK</td>
<td>N of courses in: Math</td>
</tr>
<tr>
<td>Raymond &amp; Fletcher, (2002a; 2002b); Raymond et al. (2001)</td>
<td>BWE</td>
<td>Program entry: TFA vs. All</td>
</tr>
<tr>
<td>Taddese (1997)</td>
<td>CK</td>
<td>Major in: Math</td>
</tr>
</tbody>
</table>

* Betts et al. estimates for minors are not included in this analysis
Chiang et al. double degree is not included in Figure 4 because it confounds Bachelors degree and Masters degree.
Rowan et al. not included in Figure 4 because we could not convert the data into a gain score metric.

**DATA ANALYSIS**

We have devised a novel yardstick for defining influences on student test scores. We want to see whether teachers who took more courses have students with higher achievement gains than other students. But the apparent influence of a college course will look different when student achievement is measured on a 50-point scale than when it is measured with a 10-point scale. The conventional solution to this problem is to divide all effects by the group’s standard deviation, thus creating a scale that puts all effects on a common
metric. But this approach can only work when all researchers derive their standard deviations in the same way (see Hedges, 1986 for more on this). If one researcher provides a pooled within-school estimate of variation while another provides an estimate that includes between-school variation, then the standard deviations will not convert effect sizes to a comparable metric. Moreover, even if the standard deviations were comparable, we face another problem when studies rely on national data bases and their associated tests, for these tests are not aligned with local curricula. They will likely measure some content that was not taught and fail to measure some content that was taught. Because they miss some areas the teacher has influenced, and measure some areas teachers have not tried to influence, we expect these tests to underestimate teachers’ actual instructional influence on students. This is not due to inadequacies in the teacher but to a lack of alignment between the teachers’ curriculum and the test content. Hence their effects are not comparable to effects that are based on state or local tests which are designed to reflect the local curriculum. In fact, this under-estimation is apparent in the average annual gains shown in national data bases, which are often just one or two points for the entire school year.

Our solution to this “yardstick” problems is to define all outcomes relative to the average amount that students gain during a given school year. That is, instead of standardizing effects by dividing them with the standard deviation, we standardize by dividing them with the average annual gain for the population. For instance, if majoring in mathematics adds 1 point to students’ achievement scores, and if students typically gain 3 points during the year, then we would say that the effect of the major was equivalent to 1/3, or 33 percent, of students’ annual gain. Our estimates of effects, then, are based on the following equation:

\[
b_i \frac{\bar{Y}_{\text{post}} - \bar{Y}_{\text{pre}}}{\bar{Y}_{\text{post}} - \bar{Y}_{\text{pre}}}
\]

where

\(b_i\) is the unstandardized regression slope for a particular qualification taken from a particular sample; and \(\bar{Y}_{\text{post}} - \bar{Y}_{\text{pre}}\) is the gain in mathematics achievement for the population from which the sample was taken.

This strategy offers certain advantages over other standardizing strategies. We do not have to eliminate studies because they calculate their standard deviations in different ways, for instance. And, at least for local data bases, we do not have to worry about how closely student tests are aligned to local curricula because these are equalized by the use of average gains as our yardstick.

FINDINGS

Findings are presented in two sections. In the first, we summarize research that examines the influence of teachers’ curricula on their students’ mathematics achievement. These studies allow us to evaluate hypotheses about the relative importance of pedagogical knowledge, content knowledge, and pedagogical content knowledge. In the second section, we examine two sets of literature that are relevant to the bright, well-educated person hypothesis. These include multiple regressions that measure the status or selectivity of the institutions that teachers attended, and a handful of studies that contrast TFA teachers with other teachers.
Evidence for the influence of college courses

When people suggest that teachers need more knowledge of, say classroom management, they really mean more knowledge of this relative to other areas of knowledge. And when people argue that teachers do not need knowledge in an area, they rarely mean that this knowledge is absolutely lacking in value, but rather that it is not as valuable as some other knowledge would be. Defined in this way, the empirical question then becomes one of identifying the value added by different parts of the teachers’ total college education, where one part might be courses in education, another might be courses in subject matter (in our case, mathematics), and a third might be courses in math education, or how to teach mathematics.

Because estimates of effects can depend statistical models—that is, on the combination of other variables the researcher measure and control in their model—it is not possible to directly compare one study’s estimate of effects with another. We solve this problem in two ways. First, we examine studies that test more than one hypothesis within a single statistical model. These studies allow a direct comparison of different course content under the same methodological conditions. Our second strategy arrays findings from studies conducted under more varying statistical conditions. We have already made these studies somewhat more comparable by requiring that they all fall within our best-evidence boundaries. We also increase comparability by examining all effects relative to student achievement on that particular test.

Within-study comparisons

Here we examine two studies that meet our best-evidence standards and also test multiple hypotheses within a single study. The first (Monk, 1994) examines secondary mathematics teachers and the second (Harris and Sass, 2006) examines all grade level teachers.

Monk used multiple regression, a statistical technique that allows researchers to estimate the contributions of several different potential influences within a single equation. Monk’s equations included student background variables as well as several measures of teachers’ educational backgrounds. He estimated the influence of individual courses teachers took as well as the influence of majoring in mathematics. Since Monk studied both sophomores and juniors, we review two equations, one for each group. Figure 61.1 displays his findings.

Each hatch-mark in Figure 61.1 represents the amount that students’ test scores increased (or decreased) with each additional course their teachers took in math or math education, assuming that everything else that was measured is held constant. Recall that the “amount” is measured relative to the average gain for all students in the population, so it reads as a percentage of that gain. So, for example, assuming no change in students’ backgrounds and assuming no other measured differences among teachers, the top line indicates that, among teachers of sophomores, each additional course taken in math education is associated with an increase in student achievement equivalent to about 16 percent of sophomores’ average achievement gains.

The general pattern, for both sophomores and juniors, is to see the largest benefits in mathematics education courses, somewhat smaller benefits in mathematics courses, and a negative effect for majoring in mathematics. In this sample of mathematics teachers, the average teacher took about eight courses in mathematics but only two in mathematics education.

Monk suspected that there might be a pattern of diminishing returns as teachers took...
more and more courses in mathematics, and tested this idea by distinguishing early course-taking from more advanced course-taking. He found that the benefits of courses in mathematics tapered off after about five courses. In Figure 61.1, the large positive effect of teachers’ college math courses for juniors represents added benefit for each of the first five courses teachers took, while the smaller number (8 percent) represents the additional benefit for each course taken beyond the first five.

The most surprising finding here is that, at both grade levels, when teachers majored in mathematics, their students actually gained less than students whose teachers did not major in mathematics—about 8 percent less for each group of students. This apparent negative result is a surprise since most of the debate about teachers’ curriculum focuses on the value of courses in education, not on courses in content knowledge. Even Monk’s diminishing returns hypothesis cannot explain this outcome.

Here we face one of the difficulties of squaring research findings with our common sense. It is hard to imagine that additional study in any topic could actually make someone a less effective teacher. The implication would be that these courses are providing harmful knowledge. We seek an alternative explanation for this pattern. Many explanations are possible, including possible flaws in the statistical model, but the one that seems most probable to us is that teachers who chose to take more courses in mathematics were already different from other college students even before they took the courses, and the difference is one that leads to less effective teaching. Perhaps college students who choose to major in mathematics have different personalities, values or beliefs that render them less effective as teachers, and it is these initial differences that explain the apparent negative effects of additional courses, not the courses themselves. This pattern of findings reminds us of the two important interpretive difficulties in this type of research: the effects we see in these studies must be interpreted in the context of the entire statistical model, and that these measures of course-taking reflect both the knowledge gained from the courses and the teachers’ original interests and dispositions that motivated her to take these courses in the first place.
Now let’s look at the Harris and Sass (2006) study. This study is distinctive in part because it offers an unusually large collection of estimates of effects and because it is an unusually high-quality study. These authors use hierarchical linear modeling, an analytic strategy that better accommodates the nested organization of teachers and students in schools, and they also used fixed effects, a strategy that better accommodates hidden differences (such as subtle personality differences or school climate differences) among teachers and schools that might interfere with interpretations of findings. The sample for this study is limited to a single state, however this (Florida) provides great detail regarding teachers’ college coursework and the data are available for elementary, middle and secondary teachers.

Figure 61.2 shows the array of effects that Harris and Sass reported. It can be read in the same way as Figure 61.1, except that these authors examined course credits rather than courses per se. Each hatch mark represents the effect of an additional credit of study in a particular knowledge domain, measured as a percentage of students’ one-year achievement gains, again assuming all other variables in the model are held constant. The figure shows a large number of estimates because these authors were able to include multiple grade levels and to distinguish specific course content—for instance, they could separate out courses in one aspect of education versus another. Like Figure 61.1, this one suggests a great variety in size and direction of effects. As we argued when reviewing the Monk study, we are not inclined to hypothesize that knowledge itself can be harmful, and therefore suspect that the apparent negative effects shown in this chart reflect either a statistical anomaly or a pre-existing difference among teachers who chose to take different patterns of courses.

Looking first at elementary teachers, Figure 61.2 suggests that credits in education and in math education are somewhat more beneficial than credits in mathematics per se (Again we assume that negative effects do not really mean that the additional knowledge has harmful effects). However, when we examine the patterns for middle and secondary teachers, the effects are far more various and difficult to interpret. All three domains of

![Diagram of Figure 61.2](image)

*Figure 61.2* Estimates of the relationship between course credits teachers earned in college and their students’ achievement gains in mathematics (from Harris and Sass, 2006).
knowledge show both positive and negative effects. While the math education courses tend to balance between positive and negative effects, education courses strongly lean toward negative and mathematics courses also lean slightly toward negative.

Harris and Sass did not test for the possibility of diminishing returns for courses taken in different content areas. However we may still imagine that people with different a priori inclinations tended to take different patterns of courses and these prior differences could account for the patterns we see here. For instance, since the estimates within education refer to specific content (class management vs. instructional strategies, etc.) it is possible that teachers who chose these courses had prior personalities or beliefs that render them relatively more or less successful as teachers, independent of the effect of the credits per se.

All best-evidence studies

Our second approach to reviewing studies of teachers’ educational backgrounds gathers together all the research findings that examine courses into a single graph. Figure 61.3 summarizes all the best evidence regarding the number of courses teachers took (it does not include the Harris and Sass data since that examined course credits rather than courses per se), and Figure 61.4 summarizes all the best evidence regarding intensive study in a particular area. In these figures, each hatch mark represents one estimate from one study. Together, though, many studies based on many different populations are represented.

The most important reason for examining all of the studies in a single graph is to search for broad patterns of effects that may rise above the variations in statistical models and other study details. Such a pattern might give us an idea of the true relationship between teachers’ educational backgrounds and their students’ achievement, one that is not an artifact of the individual studies.

Figure 61.3 still reveals quite a variety of outcomes, though they do lean toward positive at all grade levels and subjects. However, with the exception of Monk’s findings,
most effects are relatively small. Since these hatch marks represent the effect of individual courses, we might expect small effects which could accumulate as teachers took more courses. As we saw in the Harris study, there does seem to be more dispersion in effects of courses for secondary teachers than for elementary.

Figure 61.4 summarizes all estimates of the effect of obtaining a full major in either subject, or of obtaining an advanced degree in either subject. Since these hatch-marks represent the effects of extensive study in one of these domains, we might expect them to be substantially larger than those of individual courses summarized in Figure 61.3. Indeed, most of these effects are larger than those in Figure 61.3, but not by much. Moreover, many are negative as well as large.

No researcher other than Monk has tested for the possibility of diminishing returns for additional courses in a given subject but such a relationship might help us understand this pattern of outcomes. For example, among elementary teachers, majoring in either education or mathematics is apparently detrimental to student achievement, a finding that contradicts the more generally positive pattern shown in Figure 61.3 of effects of individual courses. Moreover, we found another negative effect of majoring in mathematics (from Rowan et al., 2002) which we were unable to convert to our metric and so are not displaying here. But the contradiction between the apparent benefits of individual courses and the apparent detriment of extensive study is at least partially reconciled if we assume that there are diminishing returns as teachers accumulate more and more courses in any one domain. Additional courses benefit teachers up to a specific point, but add little or no benefit after that point. We might expect to see the effects of a college major flipping from positive to negative as different researchers use different statistical models, none of which can accommodate a curvilinear relationship. Or we might assume that the people who choose to major in either subject are somehow different from those who take a few courses in each area but major in neither.

![Diagram](image-url)

**Figure 61.4** Estimates of the relationship between teachers’ college majors and fields of advanced study and their students’ achievement gains in mathematics, all studies, 1980–2006.
Only at the level of advanced degrees do we begin to see a clear difference across domains of knowledge. Though there are fewer estimates of the effect of advanced degrees, the pattern in education clearly differs from that in mathematics. Both estimates of the effect of an advanced degree in education are negative and all three estimates of the effect of an advanced degree in mathematics are positive. We need to add one caveat to this observation: all of these estimates of the effect of advanced degrees come from a single data base, the National Educational Longitudinal Study of 1988. It is possible that these results are an anomaly unique to this data base. Ludwig & Bassie (1999) are convinced that the pretest used in this data base is not adequately controlling for non-random assignment of students to their teachers. However, if the finding is real and generalizable, it raises an important question about the value of obtaining a masters degree in education. We offer two speculations here. First, since most school districts provide salary benefits to teachers who earn advanced degrees, practicing teachers have an incentive to seek them out, and education programs may try to oblige them by offering degrees that are not very rigorous or beneficial to their teaching practice. Hence the degrees themselves may be relatively weak. Second, it is possible that the type of person who chooses to obtain an advanced degree in education differs from other teachers in being less effective to begin with and that the degree does nothing to improve their effectiveness.

Summary

In this section, we have examined evidence regarding three hypotheses about the kind of knowledge teachers need. We are interested in the relative merits of pedagogical knowledge, pedagogical content knowledge, and content knowledge. We used coursework in education to test the hypothesis that pedagogical knowledge makes a difference, coursework in mathematics education to test the hypothesis that pedagogical content knowledge makes a difference, and coursework in mathematics to test the hypothesis that content knowledge makes a difference.

Despite variability among study findings, we see some broad patterns. First, courses in all three domains of knowledge appear capable of improving elementary teachers' effectiveness. For elementary teachers, the modal benefit of additional courses in education and in mathematics education appears to be around 0–3 percent of students' annual achievement gains, with courses in mathematics adding somewhat more than that. Second, estimates are far more variable among middle and secondary teachers than among elementary teachers, whether we are looking at individual courses or intensive study in a domain. This may be because there are more anomalies in secondary-level data bases, perhaps more confounding variables in secondary studies due to local practices of assigning students and teachers. Still we can make out some trends. In particular, the differences between Figures 61.3 and 61.4, suggest that courses in all three domains have diminishing returns, such that the effects of majors, in either education or mathematics, are highly variable but on average very close to zero. Finally, we see a clear difference between mathematics and education in the benefits of advanced degrees, and suspect this difference may be due to both the differences among teachers who choose to seek out these degrees as well as differences in the benefits of the degrees themselves.

Throughout this text we have considered a variety of hypotheses to account for the patterns we see. The hypothesis of interest, of course, is that knowledge in these domains benefits teaching. But the presence of negative effects raises the question of whether knowledge can be detrimental to teaching. We are inclined to believe that these apparent negative effects are not due to the courses themselves, but rather to either pre-existing
differences among people who chose these different curricular paths, or to statistical anomalies, perhaps resulting from the presence of curvilinear relationships or multicollinearity on the statistical models employed.

Evidence for the bright, well-educated person hypothesis

We rely on two very different types of studies to test the bright well-educated person hypothesis. One group of studies examines the influence of the status, or selectivity, of teachers’ alma maters on their current students’ achievement. The other examines the effectiveness of teachers who were recruited into the profession via the Teach for America (TFA) program, a program that explicitly recruits people presumed to be bright and well educated, but who have not necessarily studied teaching.

Institutional status

Two studies examined indicators of institutional status, or selectivity, in their attempts to account for teacher effectiveness. They differ in how they define and measure institutional status, however, so each must be examined separately. Figure 61.5 and 61.6 show the pattern of changes in students’ mathematics achievement that are associated with the status of their teachers’ alma maters.

The first study (Clotfelter et al., 2004a, 2004b) is the only study in our data base to explicitly recognize that teachers and students may already be matched by their social class and qualifications even before any data are collected. These authors checked individual schools to see how students were assigned to teachers. They found that many schools had assignment policies that placed less advantaged students with different teachers than more advantaged students. In Figure 61.5, we see the difference in the apparent benefit of the status of teachers’ alma maters, as classified by Barrons’ College Ranks, when examined with the full sample and again when examined only in those schools that relied on random assignment practices. The figure suggests that higher alma mater status is indeed associated with higher student achievement in the full sample,
where students and teachers are matched, but that the opposite pattern appears when the study is restricted to schools that randomly assign students to teachers. In schools where students are not matched to their teachers, teachers from the most prestigious alma maters appear to be less able to foster learning in their fifth grade students.5

Figure 61.6 summarizes findings from the second study (Aaronson et al., 2003), which focuses on ninth grade mathematics teachers and defines institutional status with a scale based on US News and World Reports classifications. We might expect affinity assignments to be even more prevalent in secondary schools than in elementary schools, since secondary schools tend to engage in more tracking, but Aaronson and colleagues tested for this possibility by comparing their observed classroom assignments with a variety of simulated assignment possibilities. Based on these analyses, the authors argue that their data did not suggest systematic classroom sorting. With respect to outcomes, the pattern suggests that students’ mathematics achievement tests scores were usually higher when their teachers had attended higher-status institutions. But the pattern is also uneven and suggests that only the level 5 institutions stand clearly apart from the others. This may be due to unusually small samples in some of the status categories. Since the first study was conducted in elementary schools and this one in secondary schools, it is possible that institutional status has a greater impact in secondary school than in elementary school.

Teach for America recruits

The second type of study that is relevant to the BWE hypothesis examines teachers who were recruited into the profession by Teach for America (TFA). On its website, TFA defines its recruitment strategy as follows:

Each year, Teach For America launches an aggressive effort to recruit the most outstanding graduating college seniors and recent college graduates—people who will be the future leaders in fields such as business, medicine, politics, law, journalism, education, and social policy. We seek a diverse group—socio-economically, racially,
ethnically, politically, and in every other respect. We seek leaders who can describe significant past achievements and who operate with an exceptional level of personal responsibility for outcomes. Because our corps members face such tremendous challenges, we seek applicants who have demonstrated determination and persistence when confronted with obstacles in the past. Lastly, we seek people with the specific skills—from critical thinking to organizational ability—that we have seen characterize our most successful teachers.


TFA does not assume its recruits will take up teaching for their entire careers, but does ask for a two-year commitment and most recruits apparently honor that commitment.

Once admitted into the program, TFA enrolls candidates in a 5-week summer institute that requires them to teach summer school students while concurrently taking courses in such topics as classroom management, learning theory, student diversity and so forth. To test the benefits of this program, researchers usually contrast TFA recruits with other teachers.

There is considerable muddle about what one is learning from such a comparison. Some researchers think that this comparison informs them about the value of pedagogical knowledge, assuming that the comparison group of traditionally-certified teachers has taken courses in pedagogy that TFA recruits have not taken. But TFA teachers are not completely lacking in pedagogical knowledge because TFA does provide some of this content to its teachers. In addition, TFA candidates must also meet state requirements for certification and many states require that teachers who enter the field without a traditional certificate must work toward obtaining one during their first few years of practice. In these cases, differences in pedagogical knowledge between TFA recruits may be small at the outset and may disappear completely as TFA recruits take the courses required for traditional certificate.

There is another reason why these comparisons don’t inform us about pedagogical knowledge: in the studies we review here, most teachers in the “other” category also lack a full certificate. The non-TFA teachers in these studies represent an eclectic mix of emergency-certified, alternatively-certified and traditionally-certified teachers. For all these reasons, then, contrasts of TFA teachers with other teachers are not very informative about the value of pedagogical knowledge, traditional certification, or any other aspect of teachers’ college course work. We cannot be sure how much pedagogical knowledge either group actually has.

But the contrast does inform us about the value of different recruitment strategies, since TFA focuses its recruitment on more selective institutions. So an examination of TFA graduates offers us a good opportunity to test the bright, well-educated person hypothesis in situations where pedagogical knowledge may not be substantially different.

We found three studies that meet our evidence standards and that examine the influence of TFA recruits on student math achievement scores. In each of these studies, the number of TFA teachers available for study was relatively small. The first two studies were both conducted in Houston, Texas, using data that belonged to the school district. These two studies both used multiple cohorts of teachers and students, and were not able to control students’ classroom assignments. They both used multiple regression to try to statistically adjust for other differences among students.

The first Houston study was conducted by Raymond and Fletcher (Raymond & Fletcher, 2002a, 2002b; Raymond et al., 2001). These authors used 1996–2000 data from upper elementary and middle school grade levels. During this period, Houston hired between 350 and 420 new teachers each year, but only about 20–25 per year, or less
than 10 percent, were TFA teachers. The researchers note that TFA teachers were placed in schools with higher percentages of Latinos and with higher percentages of students receiving free or reduced-price lunches. Test scores were also lower in the schools where TFA teachers taught.

The second Houston study (Darling-Hammond et al., 2005a) was explicitly designed to replicate and extend the first by paying more attention to the specific credentials of the comparison group. These authors found over 100 different certification categories in Houston’s data base and, after studying the state’s code, reduced these to seven categories: standard, alternative, emergency/temporary, certified out of field, certified no test, uncertified, and unknown. Because we are interested in the TFA as a means of testing a recruitment strategy, we do not report findings from this study that address the certification value of the TFA program. 7

The third study of TFA recruits was an experiment in which students within each grade level and school were randomly assigned to either a TFA teacher or another teacher (Decker et al., 2004). Because experiments entail random assignments of students to programs (in this case, to teachers with different kinds of credentials) they remove the confounding caused by affinity assignments that tends to occur in natural settings and they increase the chances of seeing an unbiased estimate of the effect of the program of interest.

Because the Decker et al. study is an experiment, it is stronger than the other two, but there are certain problems it shares with the other studies. In all three studies, the number of TFA teachers is rather small, so that researchers had to combine grade levels to obtain reasonable group sizes. And in all three studies, the non-TFA teachers represented an eclectic mix of credentials. For example, in their comparison group of novice teachers, Decker et al. found that 31 percent had a full traditional certificate, 28 percent had a temporary certificate, and another 25 percent had emergency certificates. Hence the comparison is not a good test of alternative credentials. However, the contrast remains a reasonable test of the bright, well-educated person hypothesis, since the main distinction between these two groups lies in TFAs recruitment strategies. Less than 4 percent of the novices in the Decker et al.’s “all other” group graduated from a competitive college or university, whereas 70 percent of TFA teachers graduated from such schools.

Figure 61.7 shows the effects of TFA recruits relative to other teachers in each of these three studies. Each hatch mark represents a particular estimate of the influence of TFA recruits on their pupils’ mathematics achievement. They are all positive, ranging from 3 percent of annual growth to 21 percent. These findings are encouraging for the bright, well-educated person hypothesis, in part because we are more confident that the normal processes of differential migration and affinity matching have been disrupted in one case.
by an experiment and in the other two by the TFA recruitment and placement strategy. So these findings are more persuasive than those about institutional status that were displayed in Figures 61.5 and 61.6. However, the most persuasive findings, those from Decker et al., also yielded the smallest effect.

**Summary**

Evidence regarding the bright, well-educated person hypothesis comes both from studies of the institutional status of teachers’ alma maters and from studies of TFA, which can be considered as more of a recruitment strategy than a different approach to certification. Studies in the first group are more susceptible to biases associated with affinity assignments, but those in the second group correct for these biases by systematically drawing teachers from elite institutions and placing them in schools with less-qualified students. The second group, however, is limited to very small samples of bright, well-educated teachers which are then compared with large, amorphous samples of teachers with highly variable backgrounds. If we can assume that more selective institutions tend to produce brighter, more well-educated people, then both sets of studies provide some support for the hypothesis that bright, well-educated people do help students learn more.

For those who argue for the BWE hypothesis as an alternative to pedagogical knowledge, this evidence is reassuring but not entirely satisfying. The strongest study in our set, the experiment, suggests that bright well-educated people add about 3 or 4 percent to students’ average annual achievement gains, an amount roughly equal to the benefit of a single course in education or in mathematics education. The comparability of findings across hypotheses suggest that casting the question as one of either/or will not lead to an optimal approach to teacher education.

**DISCUSSION**

Since teachers are in the business of educating students, we want to believe that their own education is relevant to their effectiveness. Many researchers have tried to test the effects of teachers’ educational backgrounds on their students’ achievement and they have had to solve a wide range of methodological problems to do so. In this synthesis we have focused only on relatively high-quality studies and we have rendered their findings comparable by using average population gains in achievement as a common yardstick. Based on these studies, we suggest that individual courses do indeed add to elementary teachers’ effectiveness, and that this is true for all three domains of knowledge—mathematics, education, and mathematics education. Secondary students also benefit when their teachers have taken additional courses in mathematics and in mathematics education. However, there is reason to believe that additional courses eventually have diminishing returns, so that the early courses teachers take would have a relatively larger impact, but more advanced courses would add little or no further benefit. This may explain the wide variability in apparent benefits of majoring in either mathematics or education. Majoring in either subject appears to have negative effects on elementary teachers’ effectiveness and also on secondary teachers’ effectiveness, though the secondary estimates are far more variable.

There is also evidence that teachers with different personal qualities may have differential effectiveness. For example, students learn more from teachers who graduate from more selective institutions. We base this conclusion mainly on TFA studies, since the two regression studies of institutional status yielded inconsistent findings. The effects of
TFA’s selective recruitment are somewhat larger than those of individual courses teachers take across all colleges, but they do appear to be consistently positive.

Finally, there appears to be a negative effect associated with obtaining an advanced degree in education, and a positive effect associated with obtaining an advanced degree in mathematics. We suspect that this pattern may also reflect differences in the personal qualities of teachers who seek out these degrees, which are not required for an initial teaching certificate, but for which there are often financial incentives to pursue.

Particularly surprising here is the difference between the observed patterns of effects and the conventional wisdom about teacher knowledge. Among the four hypotheses that motivated this analysis, the most contentious is pedagogical knowledge. Teacher educators believe they have meritorious and important content that teachers need to know, but skeptics fear that these courses are rarely beneficial and that teachers would be better educated if the space currently reserved for education courses were released. The least contentious hypothesis is that content knowledge is essential: virtually all advocates support the importance of content knowledge. Yet the differences we see in these data do not suggest such a differential conclusion. Data in Figure 61.1 suggests that mathematics education is more beneficial than mathematics per se, and data in Figure 61.2 suggests virtual comparability across the content domains, with mixed results in all three domains and at both elementary and secondary level. Data in Figure 61.3 shows a modest benefits to courses in all content domains, while data in Figure 61.4 shows marked variability in estimates of the benefits of majoring in either education or in mathematics. It is not until we get to the level of advanced degree that we see a clear difference between education and mathematics, and these differences are based on a single data base.

Taken as a whole, our evidence does not suggest that education courses are harmful except at the graduate level. Indeed our evidence suggests that, in moderation, education courses are beneficial. At the same time, our evidence also does not suggest that mathematics courses are entirely beneficial. Instead, it suggests that that there may also be a limit to the number of courses in mathematics that will benefit teachers. Perhaps a good next step for researchers would be to consider non-linear models when testing these hypotheses, and to examine their models more closely for aberrations that might be influencing outcomes.

We suggested at the outset that, although most advocates for coursework in education are also in favor of courses in other domains, and are in favor of bright, well-educated teachers, whereas advocates for these other hypotheses tend to be quite skeptical of education coursework. In an adversarial setting, it will be tempting to dismiss equivocal findings about one domain while embracing equivocal findings about another. A better response to the data would be to begin a program of research that more closely examines the content actually taught in various courses and to generate more nuanced hypotheses about what it takes to learn to teach. Both mathematics and teacher education departments could benefit from a closer examination of the content they offer to prospective teachers, with an eye toward its relevance. Mathematics departments might want to examine the relevance of advanced mathematics coursework for teaching. Teacher education departments certainly need to examine the content they provide in their advanced degree courses.

NOTES

1 Work for this paper was supported by a grant from the U.S. Department of Education, Institute for Educational Sciences, and by the National Science Foundation, Program on Research, Evaluation
and Communication. Responsibility for the content and quality of the paper reside solely with the authors.

2 Though we use the term “value added” in our title, we use it in a more general sense than it is sometimes used. In the literature, this phrase sometimes refers to a very specific analytic approach (Kupermintz, 2002; Sanders & Horn, 1994) and sometimes to the general approach of using pretests to control for prior achievement in a statistical model (Cunningham & Stone, 2005). We use it here in the more general sense.

3 We eventually expanded our list of qualifications to include some things that typically are acquired after teachers obtain full-time teaching positions because these things may be relevant in district hiring or salary decisions. These include, for instance, teachers’ years of experience, whether or not they possess an advanced degree, whether they had been certified by the National Board of Professional Teaching Standards, and how they responded to a commercial hiring interview such as the Teacher Perceiver Interview. These later additions have no bearing on the present study.

4 There were a few cases in which authors did not provide this information. When they did not, we sometimes found relevant information in another research report that used the same data base and sometimes from a web site that presented population statistics. The general goal was to find a population average pre- and post-test score that could be used as a benchmark for examining the size of teachers’ effects on students.

5 Since these authors did not provide information on average students’ change during the year, we present their findings using raw scores. We are unable to say whether or not these differences are substantial relative to average gains.

6 Another recent study, by Laczko-Kerr (2002a; 2002b; Laczko and Berliner, 2001; Laczko-Kerr and Berliner, 2002) was rejected because it did not employ a pretest.

7 The Darling-Hammond et al. study actually includes effects as measured by several different tests, and the authors argue that, in the Houston situation, the other tests may be better indicators of student learning. However, we prefer the TAAS test because it is the assessment that is aligned with the state curriculum and that teachers are accountable for. Since, in all cases, we take only one model per study sample, to avoid data dependency among effects, this is the only estimate we take from that study.

REFERENCES


* Included in Table 3.1.