Assessing the Effectiveness of Teachers from Different Pathways: Issues and Results

*Marsha Ing and Susanna Loeb*

Teachers enter the classroom through a range of pathways, which vary in their amount and focus of course work, their required experiences in classrooms, their recruitment and selection criteria, and their costs. While many teachers still obtain their certification through completing a four-year undergraduate program or a one- or two-year master's program, others enter through a variety of other pathways, including early-entry programs with very little course work on teaching or supervised experiences in classrooms prior to teaching. Information on the effectiveness of teachers from different pathways can help to improve state policies governing preparation requirements, the design of preparation programs, and school and district teacher-selection and -placement policies. Yet, until recently, research examining the relationship between teacher preparation and teacher effective-
ness was quite limited. This chapter summarizes the relatively new research on this topic, highlighting factors to consider when estimating effectiveness. It concludes that early-entry programs have shown positive effects for some subject areas and grade levels and provides useful information about how better to structure preservice preparation requirements and practices.

Teacher effectiveness is multidimensional. Teachers can be effective at improving the learning of students in one area of the curriculum or another; they can be effective at promoting student self-esteem, motivation, or engagement. Measuring teacher effectiveness is a complex task. Ideally, we would like to see how students progress across various dimensions during the time they are exposed to a particular teacher. However, in practice we rarely have measures of student progress in any one of these dimensions over time; and, even when such measures are available, it is difficult to distinguish the contribution of a teacher from other factors such as home life, peers, and school climate, all of which also affect student progress. Understanding the effectiveness of specific teacher pathways is also quite challenging. If teachers from each pathway were randomly distributed across and within schools, then we would not have to worry about these other factors when assessing the differential contribution of teachers from each pathway; but, they are not. Some pathways lead teachers to certain types of schools and students while other pathways lead to other types of schools and students. As such, assessments of the effectiveness of pathways to teaching are best viewed according to the dimension of effectiveness they address, the extent to which they accurately capture a more-complete measure of effectiveness, and the extent to which they convincingly distinguish teacher pathway effects from other factors contributing to student progress.

This chapter focuses on measures of teacher effectiveness derived from student performance gains on standardized tests. As discussed
further below, there are drawbacks to this approach. An alternative would be to analyze how pathways into teaching affect teacher behaviors, such as instructional practices and career decisions, instead of student outcomes. One benefit of this approach is that it eliminates the need to match teachers to the students they teach. Of course, it does have the disadvantage of not actually measuring student progress; the link between teacher behaviors and students is established separately. A second alternative would be to study student progress on measures other than test performance. Unfortunately, the research in this area is sparse.

Even when analyzing only student test performance, accurate estimation of pathway effects is complicated. A naive estimation would compare average student performance of teachers from one pathway into teaching with average student performance of teachers from another pathway into teaching. The temptation then arises to conclude that one pathway to teaching is more effective than another because it has a greater average student performance. Such a conclusion would be premature because, as discussed above, teachers are not randomly assigned to classrooms. Because of systematic sorting, any reliable analysis will need to separate the effect of specific pathways from other factors affecting students, including home environment and school climate outside of the classroom.

Clear definitions of the pathways are also important. Often teachers’ preparation experiences within the same pathway differ dramatically, while, at the same time, there is substantial overlap in teachers’ preparation experiences across pathways. How pathways differ can vary by grade and by subject-area specialization. It is difficult to derive implications for reforms if we do not understand the substance and details of different pathways. In addition, there is selection of teachers into pathways. Some pathways, for example, attract teachers with particularly strong academic skills. Separating the true benefit of the pathway experiences from pathway recruit-
ment and selection differences can help to clarify the implications for reform decisions.

The results of research on teacher pathway effects are mixed. There is some evidence to suggest that there are positive effects from particular early-entry programs, especially in math; other studies find little difference in teacher effectiveness across pathways. In what follows, this chapter articulates the importance of clearly defining pathways and then discusses the potential limitations of using students' standardized test performance as a measure of the value added by teachers. It then summarizes research on student achievement for teachers from different pathways, discusses measurement and methodological issues that might account for differences in the literature, and provides suggestions for future research on teacher pathways.

DEFINING PATHWAYS

One difficulty in assessing the effects of pathways on student learning comes simply from the difficulty in distinguishing among pathways.\(^1\) Categorizations such as traditional or alternative rarely capture consistent differences in teacher experiences across pathways.\(^2\) "Alternative" programs can range from fifth-year master's degrees to one-year internships to highly selective six-week summer programs to nonselective summer programs. "Traditional" pathways can range from undergraduate programs to fifth-year programs to master's programs. Humphrey and Wechsler found dramatic variability in structure and content across seven programs that were all considered "alternative."\(^3\) As an example, preservice field-based experience ranged from few opportunities during the summer for New York City Teaching Fellows to yearlong opportunities for those in the Teacher Education Institute program in California. If we see that one pathway produces teachers who, on
average, are more effective than teachers from another pathway, we will have difficulty knowing how to use that information unless pathways are clearly defined.

Similarly, loosely defined pathways often overlap. Some programs that are typically classified as alternative share many experiences and requirements with programs that are typically classified as traditional. For example, some early-entry programs are closely affiliated with institutions of higher education that also run more-traditional teacher-education programs (see programs in Baltimore and New York City). These institutes often aid in designing and teaching the course work for the early-entry programs, and, as a result, the course work can be quite similar to that used in their traditional programs. This overlap is important to consider when interpreting the results of studies that estimate differences across pathways. If a study finds no differences across pathways, the result could be driven by the minimal differences in experiences or other characteristics, not necessarily by the lack of an effect of particular program features, were these features to actually vary.

Given the variability within and the overlap across loosely defined pathways, research showing average differences in teacher effectiveness across pathways may tell us little about productive directions for policies and practices. We learn more from research that carefully defines pathways so that they differ as cleanly as possible in their characteristics. With these definitions, we can learn more about the effectiveness of different approaches as they relate to the recruitment, selection, and development of teachers.

Just as the pathways must be clearly defined to make results meaningful and useful, the measured effects must also be clearly categorized. One pathway or program may have a positive effect on increasing ninth-grade standardized math scores. Another may have an effect on diminishing the achievement gap among fourth graders in reading. Some pathways are more likely to be effective at
preparing teachers for particular grade levels or subject areas than others; and some pathways are likely to be more effective with certain types of students. Most studies of teachers' value-added to student achievement to date have used data on student performance in math and reading. Most, though not all, have also used data on students in elementary or middle schools. To the extent that programs or pathways excel at preparing teachers for other areas of teaching, these studies may not fully capture differences. These potential differences in effectiveness by area can help us to reconcile seemingly conflicting results in the research literature and point us in directions for future research.

LIMITATIONS OF STUDENT TEST PERFORMANCE MEASURES

Student learning is a logical metric with which to measure the effectiveness of teaching. However, available measures of student achievement are never perfect indicators of what students know or what teachers have taught. Researchers have raised general concerns about whether these tests are valid measures of the domains of knowledge that we care about, whether they reliably measure student learning, and, even if they do, whether they reliably measure the aspects of learning that teachers affect.

Our schools' goals for student learning span a range of areas: math and reading; science and history; music; art and athletics. Within each subject area there are many dimensions or domains. In math, for example, students can encounter number theory, fractions, algebra, and geometry, among other topics. English language arts includes decoding, reading comprehension, and writing. No single test or sequence of tests could cover all the material even within a single subject area in a single grade. Standards and curricular frameworks help to bound the possible scope of the material, but not nearly enough for tests to cover the full curriculum.
As such, in choosing or developing tests there are always trade-offs. Some knowledge and skills are easier to assess through standardized tests than others, and often we choose to sacrifice more difficult-to-measure elements for those that are easier. Whether these trade-offs are important enough to invalidate the tests as measures of student knowledge in the chosen domains is up for debate. It is important, however, to know what the tests do and do not measure so that this information can be used in drawing conclusions about the implications of any finding.

Part of understanding the validity of a test is knowing both the ease with which the test results can be manipulated through small changes in practice and the incentives for teachers to change their practice in order to manipulate test results. For example, if there are high stakes for teachers attached to the tests, then there will be incentives for teachers to do what they can to increase test scores. While ideally this would be through improving learning, there may be easier routes, such as cheating or teaching test-taking practices that do not improve students' knowledge in the domain nor benefit them in the long run, but may raise test scores in the short run.

Even if the tests do validly measure the domains of interest, they may not be good for measuring student learning over time. We are often not only concerned with student status at a particular point in time but also in the amount of learning or change that occurs as a result of instructional experiences. To measure student change over time requires following the performance of the same student on several occasions. In most cases, this means comparing student performance in one year with performance in subsequent years. Lissitz and Huynh point out that a major assumption of using changes in test performance is that the content across grade levels is comparable. For example, we might measure student math performance in one grade level and then again in another grade level. The assumption is that test items at these different grade lev-
els accurately measure knowledge of math and are not confounded with other skills, such as reading. If the linguistic complexity of math items at the later grades interfered with a student’s ability to demonstrate their knowledge of math, there are multiple dimensions being measured. This multidimensionality makes it difficult to tease out whether student performance is actually reflecting knowledge of a single dimension (math) or a result of multiple dimensions (math and reading).

Finally, for the tests to be useful for estimating the effects of teachers, they must not only validly measure change in knowledge over time, but they must also measure those aspects of change that could be affected by teachers. Some changes in knowledge or ability may be more a function of developmental change than of classroom learning. If the tests were to primarily reflect these aspects of learning then they would not be valid instruments for measuring teacher effects. Some researchers have argued that the tendency for studies of program and teacher effects to find greater effects in mathematics than in English language arts in elementary schools reflects this phenomenon, because learning in math is more influenced by experiences in school, while progress in learning to read is relatively more influenced by factors outside of school. The degree to which the measures are aligned with and influenced by the curriculum and instructional activities affects the validity of the test as a measure of a teacher’s value-added to student learning. The timing of testing programs may also affect the ability of research to isolate teacher effects. For example, if achievement tests are given early in the year, how much of the gain between the prior and current years should be attributed to the student’s current teacher and how much to the teacher in the previous year?

In addition to concerns about the external validity of the test measure, poor reliability of the test may also hamper its usefulness for assessing teacher and teacher-pathway effects. Even the domains
and topics specifically selected for the tests will be measured with error. This is a particular concern at the tails of the performance distribution, for the students who score either quite high or quite low. Most standardized tests aim to distinguish among students, and because many students tend to have approximately average knowledge, most tests include more questions that measure knowledge at approximately the average level. Fewer questions aim at the lower and upper levels of knowledge and, thus, there is more measurement error at these tails. Computer-adaptive tests can efficiently target questions to all levels, but most research on teacher effects to date has not used these tests. Measurement error becomes an even greater concern when we look at changes over time because the error on each test is compounded. Often we are particularly concerned with the effects of practices and policies on the lowest-scoring students, and these are just the ones for whom the test is least accurate.

In sum, tests are imperfect measures of student learning, even in the domains that they cover. A single outcome measure or set of measures will never reflect the range and dimensions of student knowledge. Tests also vary in the accuracy with which they reflect true ability in a specific domain. These difficulties are compounded when we look at changes over time and link them to teachers and try to adjust for differences in the schools and classrooms in which these teachers teach. Shulman summarized these concerns on the use of student outcome measures by saying that

... indeed, as if the standardized tests were not sufficiently questionable as sufficient reflections of achievement, the analyses of change, replete with corrections for initial differences and measurement unreliability, as well as the needed premises about equal impact of teachers across all students, are seriously deficient.

Despite these limitations, test scores do have their advantages. We are unlikely ever to have a perfect measure of student learning.
While we could forsake trying to link teachers to student learning as a result of this inherent imperfection, in doing so we are likely to lose information that could be helpful for improving practices if used thoughtfully. We do know that traditional measures of success such as teacher satisfaction and feelings of preparedness are fraught with error, and we have not yet developed accurate instruments for capturing high-quality instruction. The imperfections of the test measures should be considered when drawing implications from the research, noting which domains of learning they cover and which they do not, and assessing the extent to which the particular test used is likely to have been subject to gaming behaviors linked to high-stakes accountability. With these caveats, however, these outcome measures provide some of the only information we have on the effectiveness of different pathways into teaching and of education policies and practices, more generally.

METHODS TO ATTRIBUTE OBSERVED DIFFERENCES IN STUDENT ACHIEVEMENT TO TEACHERS

Even with clear definitions of pathways and valid and reliable measures of student achievement, researchers must design their analyses carefully in order to avoid attributing to teachers and to their pathways what are actually the effects of other factors. Teachers from some pathways teach in schools and classrooms with characteristics that are systematically different from those of teachers from other pathways. For example, in New York City, new teachers from the largest early-entry program, the Teaching Fellows, on average, taught 42 percent black students, 50 percent Hispanic students, and 89 percent students eligible for free lunch, compared with teachers from the traditional college-recommended route who taught 34 percent black students, 39 percent Hispanic students, and 76 percent free lunch–eligible students. If those differences
affect student learning, then the researcher must account for them in order for their estimates of effects not to be biased. There are several approaches to making this adjustment, and these approaches vary in their advantages and disadvantages.

**Random Assignment**

Random assignment of teachers to classrooms is the cleanest approach for addressing potential differences in the teaching context. Mathematica Policy Research, Inc. conducted such a study of pathways into teaching, randomly assigning students to Teach For America (TFA) teachers and non-TFA teachers. The goal of randomly assigning students to teachers or teachers to classrooms is to create equivalent groups of students through chance assignments. While the researcher is able to observe some student traits, there are always important characteristics that the researchers cannot observe, such as students’ enthusiasm for learning. Random assignment of adequately large numbers of teachers and students can create equivalent contexts along dimensions that include both observable and unobservable characteristics.

The main drawback of random assignment is that it is both costly and difficult to implement. The random assignment of students to particular teachers is not always feasible and successful implementation is always of great concern. As an example, students and their parents may work against random assignment, trying to get into particular classrooms with particular teachers. Because of the cost of implementation, the scope of the studies is often small and not representative of the full population of interest. Campbell and Stanley describe randomization as “not for the purpose of securing representativeness for some larger population” but “solely for the purpose of equating experimental and control groups or the several treatment groups.”
Statistical Controls and the Value-Added Methodology

Most studies of the effects of pathways into teaching are not able to implement large-scale random-assignment designs and instead rely on available data, usually administrative data from school districts or states. The researchers then use statistical adjustments to account for differences in the context of teaching across teachers from different pathways. Value-added modeling is a general classification used to describe regression-based approaches to estimating the effects of teachers or teacher characteristics on growth in student achievement. McCaffrey, Koretz, Lockwood, and Hamilton describe value-added models as “any educational achievement model that uses gain scores or regresses current scores on prior scores.” The models differ in terms of how they define value-added, the data they need, and the assumptions that they make.

Most models attempt to control for aspects that influence student achievement, such as students’ prior achievement or the demographics of the student population at the school. The motivation for including these controls is to attempt to limit factors that might explain student achievement so that the model can more accurately attribute changes in student achievement to the instructional opportunities provided by a particular teacher. For example, if teachers from one pathway systematically taught students with lower initial achievement than teachers from another pathway, if we then looked at student achievement at the end of the year without adjusting for initial achievement, we might attribute lower scores to teachers from the first pathway even if the learning was actually equal or even greater.

Value-added studies of teacher effects differ along a number of dimensions. Some studies use student achievement as the outcome measure and control for prior achievement while others model student learning as the change in test scores directly. Some include characteristics of students such as their gender, age, or race, while
others include indicator variables (or fixed effects) for each student, and by doing this, control for all time-invariant characteristics of students, even those not observed. Some include classroom characteristic measures such as the average prior test scores of students, while other studies do not. Some include school characteristics while others include indicator variables for schools, thus estimating the effects of pathways by comparing teachers within the same school. There is no consensus in the research community as yet concerning which model is optimal. Many studies try multiple specifications in order to test the robustness of their findings. That said, all reasonable studies attempt to adjust for likely differences between teachers from different pathways in the characteristics (both measured and unmeasured) of the schools and classrooms in which they work.

**TEACHER PATHWAY EFFECTS**

There are two potential mechanisms through which routes into teaching can produce effective teachers. First, through recruitment and selection, they can attract individuals who will be good teachers, irrespective of the teaching education they receive. Second, they can provide opportunities for individuals to develop their teaching skills through course work and fieldwork in classrooms. The effect of any program or route into teaching is a combination of these two mechanisms. Traditional university-based teacher-education programs have tended to focus on the second mechanism, with little emphasis on recruitment and selection. Alternative routes vary more in their emphases. The TFA program, as an example, is highly selective. It received approximately 19,000 applications for the 2006 academic year for approximately 2,400 spots. The applicant pool included 10 percent of the senior class at Spellman and Yale and 8 percent of the senior class from the California Institute of Technology.20
The effect of teachers from a given pathway is a combination, then, of both selection and opportunities to learn. Few studies to date have done a good job of separating these two mechanisms, though such separation would be useful for aiding the design of more-effective routes. If programs are successful solely through selection or recruitment, then the key to improving teaching would be to understand which selection criteria are important (for example, it could be exceptional performance on some measure of ability) and then work to attract individuals with these selected characteristics. The resources currently expended on preparation could instead be used to increase the appeal of teaching for individuals with the characteristics identified as important for successful teaching. If on the other hand, the mechanism through which a pathway is effective relies on the educational opportunities it provides, then understanding what those opportunities are and providing them to other potential teachers could be the most productive approach to improving teaching. In practice it is likely to be a combination of these two mechanisms that determine whether programs are effective.

As noted above, both alternative routes and the traditional routes to which they are compared can vary a great deal, and, as a result, the research results on the effectiveness of alternative routes relative to other routes into teaching is mixed. Much of the extant research has compared teachers from the TFA program to other teachers in their schools. TFA is a national program that recruits recent college graduates for teaching in schools that districts have had a difficult time staffing. The program requires a two-year commitment from its participants and provides a five-week summer preparation institute prior to teaching, as well as supports during teaching. The studies of TFA, as summarized below, have tended to find some positive effects of TFA teachers in math performance and little effect or negative effects in English language arts (ELA).
Studies that estimate the effects of teachers from other early-entry programs tend not to be as positive.

Decker, Mayer, and Glazerman from Mathematica Policy Research Inc. conducted a study of TFA teachers that randomly assigned teachers to students in an attempt to equate the teaching context of TFA and non-TFA teachers. They found positive effects on elementary math achievement for TFA teachers compared to teachers who did not participate in TFA. These effects held true across grade levels, geographic locations, and subgroups of students. This study did not include information about the preparation experiences of the teachers who were not trained by TFA, and thus the precise nature of the comparison is not clear.

Raymond, Fletcher, and Luque found similar results to the Decker et al. study using data on elementary school teachers from a Texas school district. Although this was not a random-assignment study, the researchers did control for student characteristics (ethnicity, free or reduced-lunch eligibility, and prior achievement), teacher's years of experience, classroom characteristics (free or reduced-lunch eligibility and prior achievement) and school characteristics (ethnicity and free or reduced-lunch eligibility). They found that over the course of a year, the students of TFA teachers gained more on tests of mathematics (but not reading) relative to students of other teachers. The authors also found less variability in the student performance of TFA teachers compared to the other teachers, although the group of other teachers came from a diverse set of routes and the study did not separate the comparison group by route. In a study using data from the same district, Darling-Hammond, Holtzman, Gatlin, and Heilig separate TFA teachers who have gained certification from those that have not and find more positive effects for the certified teachers. Uncertified teachers, both those from TFA and those from other programs, do not show as strong value-added as certified teachers in this study.
A third set of studies used data on third- through eighth-grade teachers in New York City schools. Both Kane, Rockoff, and Staiger and Boyd, Grossman, Lankford, Loeb, and Wyckoff used a variety of specifications, controlling for multiple student, classroom, and school characteristics, and including indicator variables for schools and/or students in order to adjust for differences in teaching context between TFA teachers and other teachers. The two studies used slightly different comparison groups. The Kane et al. study compared TFA teachers to teachers who obtained certification through recommendations from their teacher-education program or who obtained certification through a process of individual evaluation from the state. The Boyd et al. study compared TFA teachers to just the college-recommended teachers. In both cases, the studies found that the students of TFA teachers, on average, experienced greater gains in achievement in math and less in reading than students of other teachers. The Boyd et al. study separated students in upper elementary school (fourth and fifth grade) from those in middle school (sixth through eighth grade) and found that the positive effect of TFA teachers on student performance in math was specific to the middle school, though the negative effect in reading held for both elementary and middle school students.

Only one study to our knowledge has examined the value-added of teachers from different pathways to high school students' test-score performance. This lack of research is due largely to data limitations. High school students are taught by multiple teachers each year, and often the students do not take standardized tests in each year—or if they do, the tests are broad enough that it is difficult to link them to particular teachers. North Carolina, however, does require end-of-course exams for high school students. Using this information, Xu, Hannaway, and Taylor find that students of TFA teachers learn substantially more during the course of the year than students of teachers from traditional routes. They estimate that
the difference in effectiveness between the routes is approximately equal to twice the difference between the average first-year and average second-year teachers. There is not a large enough sample size of TFA teachers in North Carolina high schools to separate the effects by subject area and the results are an average of teachers in algebra I, algebra II, biology, chemistry, geometry, physics, physical science, and English I. The positive results for TFA teachers are driven largely by positive effects in math and science classes. The models include student indicators and thus identify the effects of the pathways by comparing changes in individual students' learning over the time they spent with teachers from different pathways.

In summary, there is some evidence that TFA teachers are at least as effective as other teachers in their schools at improving the math achievement of students. Their effectiveness appears greater at the middle and high school level than in elementary school. At the elementary level, there is evidence that they are not as effective as teachers from more-traditional university-based programs at teaching English language arts. Of course, one of the reasons TFA teachers are in schools is because those schools have trouble recruiting teachers from traditional university-based programs. Zeichner and Conklin question these conclusions about the effectiveness of TFA teachers; they argue that the research does not adequately describe how the TFA teachers were prepared or how their preparation differs from that of non-TFA teachers. They conclude that based on "the lack of information about the characteristics of the teachers, one cannot differentiate the effects of teacher characteristics from those of teacher education programs." As Zeichner and Conklin suggest, these results do not necessarily indicate that TFA provides prospective teachers with uniquely effective preparation opportunities; in fact, the positive TFA effects may largely be due to their emphasis on recruitment or selection rather than preparation per se. Nonetheless, as mentioned above, the combination
of selection and preparation characterizes any route into teaching, and thus these results are as accurate an assessment of the effectiveness of the route available today.

While much of the research effort has focused on evaluating the relative effectiveness of TFA, there are many other alternative routes and even early-entry specific routes which offer diverse training opportunities and serve substantial numbers of prospective teachers. For example, in the 2004 school year, TFA provided 360 new teachers to New York City schools, while the New York City Teaching Fellows, a different early-entry program, provided 2,441 new teachers. Teaching Fellows is modeled on TFA in that it places substantial emphasis on recruitment and selection and provides opportunities to learn about teaching during a relatively short preparation period, often in the summer.28 It is not quite as selective as TFA, and its preparation opportunities differ. As an example, while TFA requires its prospective teachers to coteach in a classroom during the summer program, Teaching Fellows are more likely to observe in the classroom of a New York City public school teacher for their field experiences, with more limited opportunities for independent teaching.29

Boyd, Grossman, Lankford, Loeb, and Wyckoff and Kane, Rockoff, and Staiger also evaluated the effectiveness of Teaching Fellows teachers. The results are somewhat similar though not as positive as those for TFA.30 Boyd, Grossman, Lankford, Loeb, and Wyckoff found that relative to the students of traditional-route teachers, the students of first-year Teaching Fellows do not improve as much in their performance on standardized tests in English language arts (ELA) in either elementary or middle school, and do not improve as much in their performance in math in elementary school. The negative effect in ELA is approximately the size of the difference in effectiveness between the average first-year teacher and the average second-year teacher in the sample. There is no discernible dif-
ference between students of teachers from these routes in middle school math. After the first year, however, Teaching Fellows do, on average, as well as teachers from college-recommending programs, except in middle school ELA. As Kane, Rockoff, and Staiger point out, while the difference between the Fellows and traditional or college-recommended teachers is evident in the first year, the variation in effectiveness as measured by the teachers’ value-added to student achievement varies much more within programs and pathways than between programs and pathways. That is, there are both very good and very poor teachers in all of the routes studied.

As might be expected, the effect of pathway appears to be larger for first-year teachers than for other teachers. Because of this, the total effect of the pathway on teaching quality in a district is a combination of the number of teachers the pathway provides and the experience of those teachers. Boyd, Grossman, Lankford, Loeb, and Wyckoff reasoned that “if one pathway consistently has higher turnover, even if its teachers do well relative to those in other pathways with the same experience, the pathway may not, on average, be providing the most effective teachers.”

To investigate this further, the authors simulated the effects of pathways accounting for different tendencies for teachers from different pathways to leave New York City teaching jobs. They found that for middle school math, students from Teaching Fellows and TFA teachers perform higher than students with teachers from other pathways regardless of how long the pathways have been operating. For elementary math and ELA, the simulation results indicated that when the pathways are just introduced in a district, students with Teaching Fellows and TFA teachers initially gain less in their test performance than students with teachers from other pathways; but, within a couple of years there is little difference across pathways. ELA middle school students from Teaching Fellows or TFA teachers tend to have a lower performance during the first year of program
operation but outperform other students by the fifth year of program operation. The results suggest that overall differences in the effectiveness of teachers across pathways are not large, once variation in attrition is accounted for, although there are meaningful differences in the effectiveness of first-year teachers.

In a smaller-scale study of student learning, Miller, McKenna, and McKenna examined the relative effectiveness of teachers from a university-initiated early-entry pathway in Georgia. According to the researchers, this program was not highly selective, unlike the TFA and the Teaching Fellows programs. Entry requirements did not include test performance or a minimum grade point average. The program included coursework over the summer that complied with Georgia Certification Standards. During their first year of teaching, students received mentoring from university faculty and public school teachers and enrolled in additional course work. The researchers created matched pairs of alternative and traditionally certified teachers who started teaching in the same year, taught the same subjects, at the same grade level, at the same school. They compared classroom performance of fifth- and sixth-grade students on reading and math assessments; the researchers did not include any covariates in their analyses because they concluded that there were no differences between the groups of students in terms of a pretest measure. The authors looked specifically at teachers with three years of teaching experience and found no difference on student performance in reading and math between the two groups of students. These results are consistent with the other available studies that find discernible effects are largest in the first year. Even if there were initial differences, these might be gone by the third year. In addition, this study was so small, including only eighteen classrooms, that it may not have been able to distinguish effects even if those effects were meaningfully large.
Studies that compare the effects of certified teachers to noncertified teachers do not provide direct evidence of differences between alternative- and traditional-route teachers because both of these routes lead to certification, but they do provide some information of whether and, if so, how routes into teaching can produce teachers who are differentially effective. Most teachers in the United States are certified. For instance, in 1999–2000, 94.4 percent of public elementary and secondary teachers were certified in their main teaching assignment. In theory, certification keeps individuals who are likely to be poor teachers out of the classroom. The evidence on the effect of certification is, however, mixed.

Recent studies in New York City and North Carolina found that students of certified teachers learned more, on average, than did students of uncertified teachers, though a similar study in Florida found no difference. The studies in New York and North Carolina found that teachers who passed their certification exam (the Liberal Arts and Science Test in New York and the Praxis II in North Carolina) showed higher student achievement in math. For example, teachers who passed the Praxis II produce on average student achievement gains that were in the range of 3 to 6 percent of a standard deviation higher (in math) than those who failed. Comparing the effect of this gain to that produced by experience, the study found that the average teacher who failed the test, were he/she allowed to teach regardless, would likely produce the same level of math achievement in his/her second or third year of teaching as a novice teacher who passed the test. The study also shows how test cutoff criteria can generate a number of false negatives (individuals who fail the test but might have been high-quality teachers) and false positives (individuals who make the cutoff might turn out to be poor teachers), calling into question the signal value of certification tests. Also worrying is evidence that raising cutoff scores
might reduce the supply and racial/ethnic diversity of the prospective teacher pool.\textsuperscript{37}

Another line of research that sheds light on the effects of different routes into teaching estimates the effects of features of preparation programs on teachers' value-added to student learning. While research in this area is new, one study in New York City finds that teachers who have opportunities to practice skills that are well connected to the work in the classroom have greater value-added in their first year of teaching. For example, programs that oversee student-teaching experiences more carefully or require a capstone project supply significantly more-effective first-year teachers. Teachers who have had the opportunity in their preparation to engage with student work (listening to a child read aloud for the purpose of assessment, planning a guided reading lesson, or analyzing student math work) also show greater student gains during their first year of teaching, as do teachers who have had the opportunity to review curriculum.\textsuperscript{38} While this work is still in its infancy, it suggests that the learning opportunities that potential teachers have during their preparation can affect their value-added to students, at least in their first year of teaching.

In summary, the research literature to date linking teacher pathways into teaching with student achievement gains is sparse. Almost all existing studies have assessed the effects of one program, albeit a large one, TFA. The results here have been mixed, with indications of positive effects of TFA teachers, particularly in math, and some indication of negative effects in elementary reading. The size of the positive effects in math range from small to relatively large, up to twice the difference between the average first- and second-year teachers at the secondary level. The estimated effects for other routes, such as the New York City Teaching Fellows, are somewhat less positive, though consistent in that the elementary reading effectiveness appears the least strong and the math effectiveness for
older students, the most strong. The variation in teacher effectiveness across teachers that went through the same pathways is larger than the average differences in teacher effectiveness between pathways. These differences, particularly the negative effects of Teaching Fellows, are often stronger in the first year of teaching than in subsequent years. Overall, the research is not rich enough to provide clear policy directions; however, it does point to the potential benefit of bringing in teachers with strong content knowledge and academic ability in the high school, while at the same time suggesting the potential importance of other types of knowledge—perhaps pedagogical knowledge for teaching reading or knowledge of child development—particularly in the elementary school.

**OTHER MEASURES OF TEACHER EFFECTS**

Student test scores are not the only measure of student progress worth considering in an analysis of teachers from different pathways. Other student outcome measures such as course-taking behaviors, motivation, high school graduation, or post–high school plans can provide further information on effectiveness. Unfortunately, there are few studies that use these alternative measures for students.

In their study of TFA teachers, Decker et al. did include non-achievement student outcomes, such as whether students were retained at grade level, attended summer school, were tardy or chronically absent, or were well behaved, in addition to student achievement outcomes. Using data from teacher reports of student behavior and school and district records and the random-assignment methodology, they find no evidence that students in TFA classrooms were more likely to be retained at grade level or attend summer school compared to students in other classrooms. There was also no difference between TFA and other teachers in school-reported absenteeism or disciplinary incidents (suspensions, expul-

179
sions). Teacher reports of absenteeism and disciplinary incidents showed some differences in student behaviors, with TFA teachers reporting higher levels of problems with physical conflict among students and greater interruptions during class to deal with student disruptions compared to non-TFA teachers.

Other studies have used measures of teacher quality not directly linked to student outcomes. For example, Miller, McKenna, and McKenna use measures of instructional practice based on observations of their sample teachers. They find no evidence of a difference between the teachers from the early-entry route and other teachers in their Georgia sample in either lesson components or pupil-teacher interactions.\textsuperscript{40} A new study of alternative-route teachers in New Jersey uses interviews with superintendents and principals to assess the quality of instruction provided by alternative-route teachers.\textsuperscript{31} It finds that these administrators are satisfied with the performance of alternative-route teachers, especially in high schools where they cite the importance of their subject-matter competency. In keeping with the test-score literature, the superintendents and principals are less positive about the effectiveness of the alternative-route teachers in elementary school, citing the importance of understanding child development. In a related study, Jelmberg surveyed principals on the instructional skills and instructional planning of traditionally certified teachers and alternatively certified teachers in New Hampshire.\textsuperscript{42} While his response rates were low, calling into question the generalizability of the results, he found that principals rated traditional-route teachers higher than teachers from the New Hampshire alternative routes on both instructional planning and instructional skill.

Teaching quality is multidimensional and there are multiple measures of quality worth exploring in studies of teachers who entered through alternative routes; however, the extant research has only limited analysis of differences across teachers based on
pathways using anything other than student performance on standardized tests. Existing research does show little difference between groups, though the differences that are there are consistent with the test-score-based research.

**IMPLICATIONS FOR RESEARCH**

The research to date on the effectiveness of different pathways into teaching is sparse, but there are a number of high-quality studies that shed light, not only on the effectiveness of specific programs, but also on the factors to consider when evaluating the effectiveness of different pathways into teaching more generally.

First, as many have pointed out, not all alternative or early-entry pathways are the same. Programs differ dramatically in their recruitment and selectivity and also in the opportunities for learning they provide. If we see a positive effect of early-entry teachers in one analysis and not in another, it may well be because the early-entry programs are quite different in the two analyses. Similarly, the comparison group, whether it comprises traditional-pathway teachers or a mixed group of teachers from “other” pathways, can vary significantly as well. Differences across studies could stem from these variations in the comparison group even if alternative-pathway teachers enter through similar routes.

Second, the relative effectiveness of teachers from different pathways can vary depending on the grade, subject area, and school characteristics of the teaching context. The research suggests, for example, that teachers from highly selective early-entry pathways are more effective in math than in reading, and more effective in upper grades than in lower grades. Even if a particular pathway appears less effective in both math and ELA, that pathway might actually focus on other areas, such as teaching special education students whose learning is less likely to be accurately measured by
the standardized exams. Understanding the programs targeted subject area is important for assessing its effectiveness. Similarly, most early-entry programs serve urban schools that have been historically difficult to staff, and thus, most of the evidence on the relative effectiveness of alternative-route teachers is specific to these types of schools. These results may not generalize to other schools.

Third, differences in effectiveness between early-entry teachers and teachers who have completed a traditional teacher-education program can differ by years of experience. As an example, studies in New York City indicate that the relatively poor performance of early-entry teachers primarily takes place during the first year of teaching. These findings imply that not only will the research be more informative if it clearly distinguishes effects by years of experience, but also that the total benefit of a pathway is likely a combination of the estimated effectiveness of its teachers and their attrition rate. If one pathway is particularly good starting in the second year, but very few of its teachers stay past the first year, then its total contribution will be far less positive than if its teachers have a greater tendency to stay beyond the first two years of teaching.

Fourth, our ability to accurately measure the effectiveness of teachers from different pathways depends on the measures of student progress that are available. Often we do not have—and it is costly to collect—reliable measures of student learning or of teaching. There are clearly many domains of learning, and we may not be able to capture the ones in which we are interested. For example, we rarely have measures of achievement gains in high school classes. Some measures may also be less sensitive to teaching quality. A measure used in one study might be highly related to the types of instructional opportunities provided to the students and might be highly sensitive to the effects of instruction, while a measure used in another study might reflect the national goals or stan-
dards that are less related to the instructional opportunities of a particular classroom. Attention to the types of measures used in these studies might help account for differences in conclusions about the effects of teachers from different pathways.

Finally, research on different pathways includes different data sources and methodological approaches. The research is clear that teachers from different pathways teach systematically different students and in systematically different schools. All reasonable studies of the effectiveness of pathways into teaching work to adjust for these differences in teaching context. Studies with random-assignment designs or those with very large sample sizes and detailed data on students are best able to make these adjustments. The results of empirical studies are more or less believable to the extent that they are convincingly able to make these adjustments. Did the random assignment work to equalize baseline characteristics of students and schools across pathways? Did the analysis compare teachers within schools or do a convincing job of controlling for differences across schools?

Aggregating research findings across studies requires consideration of each of these factors. What are the characteristics of the pathways in question and of the students in question? What are the subject areas covered and the experience levels of the teachers? How is student progress measured, and how are differences in teaching context accounted for? To date the research is not rich enough to paint a full picture of the relative benefits of different pathways along multiple dimensions of subject areas, grade level, student characteristics, and teacher experience. However, the research currently available demonstrates that keeping these distinctions in mind is important; further research will contribute to our understanding largely to the extent that it can provide insights into this variation in effectiveness.
CONCLUSIONS

Recent years have seen the emergence of a range of pathways into teaching. While many teachers still enter through traditional undergraduate programs, others enter through graduate programs or through selective or nonselective early-entry programs with five or six weeks of course work and limited classroom experiences before beginning as a full-time teacher. The development of new routes into teaching has occurred largely in the absence of information concerning which aspects of preparation are particularly important for new teachers or which characteristics and skills of individuals are particularly important for teaching. It is just this type of information that will be useful for improving the programs that we have and for developing new policies and practices for entering teachers.

The research to date shows that the alternative/traditional pathway dichotomy is not particularly useful for learning about entry pathways because there is so much variation across both alternative and traditional programs, and because there is often substantial overlap between the two groups as well. Instead, the research is more useful when it identifies the differences between specific programs and practices.

As described above, there are two potential mechanisms through which routes into teaching can produce effective teachers. They can, through recruitment and selection, attract individuals who will be good teachers. Alternatively, they can provide opportunities for individuals to develop their teaching skills through course work and supervised practice in classrooms. The effect of any program or route into teaching is a combination of these two mechanisms. Traditional university-based teacher-education programs have tended to focus on the second mechanism, with little emphasis on recruitment and selection. Alternative routes vary more in
their emphases, and some have put substantial effort into recruitment and selection.

The available research does not paint a complete picture of either optimal recruiting and selection criteria nor optimal preparation opportunities; however, for a relatively small group of studies, they do provide useful and consistent results. In particular, highly selective early-entry pathways, such as TFA, appear to provide teachers for difficult-to-staff schools who are as effective, or more effective, than other new teachers in their schools at improving the achievement of students, particularly in math and at the middle and high school level. These early-entry pathway teachers have been less successful at the elementary level and in English language arts. Less selective early-entry routes may not be as effective, though the information on their value-added to student learning is very limited. On the whole, the research suggests the importance of balancing strong recruitment and targeted selection—both of which have been overlooked by most traditional teacher-education programs—with useful opportunities to learn about teaching through course work and through supervised experiences in classrooms.