The Effect of Alternative Certification on Student Achievement: A Literature Review

Final Report

March 31, 2003

Neil S. Seftor
Daniel P. Mayer

Submitted to:
U.S. Department of Education
555 New Jersey Ave., NW
Room Number 308D
Washington, DC20208

Submitted by:
Mathematica Policy Research, Inc.
50 Church Street, 4th Floor
Cambridge, MA 02138
Telephone: (617) 491-7900
Facsimile: (617) 491-8044

Project Officer:
Elizabeth Warner

Project Director:
Daniel Mayer
The past 20 years have seen a marked shift in the method of training teachers in the United States. In just two decades, the number of states offering alternative certification routes for prospective teachers increased from a handful to 45. By some estimates, about one-third of any given year’s newly hired teachers now come through alternative certification programs (Feistritzer and Chester 2002). However, the design and implementation of various traditional and alternative training programs evidence considerable variation, and research into the programs yields little consensus as to the best way to train prospective teachers. Given the variation in alternative certification programs, it is difficult to resolve issues of training effectiveness solely by relying on studies focused on the general question of whether alternative certification meets its promise. In this review, we examine studies to learn whether we can conclude what specific types and characteristics of alternative certification programs might prove effective.

As a starting point, we examine the studies reviewed by Wilson et al. (2001) in a report for the U.S. Department of Education (ED) entitled “Teacher Preparation Research: Current Knowledge, Gaps, and Recommendations,” which presents a comprehensive review of peer-reviewed journal articles on teacher preparation.1 We have also searched for reports prepared for government agencies and foundations, studies published in scientific journals following release of the Wilson article, and recently launched studies to supplement this list. Our intent was to

1 For inclusion in the Wilson et al. (2001) review, an analysis had to be a study of U.S. teacher education, directly relevant to the questions posed by the Department of Education, and published in a scientific journal within the past two decades. In addition, the authors required the study to be empirical (offering evidence rather than theory or opinion) and rigorous (meeting generally accepted standards in the relevant research areas). To qualify as rigorous, experimental and quasi-experimental studies must have used random assignment or some form of matching for entering characteristics while multiple regression studies must have controlled for relevant differences among the groups, other than the characteristic under study.
focus on studies that compare a well-specified alternative certification program to a well-specified traditional certification program as well as on those that compare standard certification to emergency, temporary, or provisional certification. Both of these comparisons would allow for a more precise examination of particular routes to certification, although, as shown below, the literature does not allow for such a fine-grained comparison.

Alternative certification is a shorthand way to refer to any of the routes—other than the traditional route—that a prospective teacher may follow to enter the teaching profession and become a “fully certified” teacher.\(^2\) We use interchangeably the terms “alternative teacher certification programs” and “programs that provide an alternative route to teacher certification.” The distinction is that alternative route programs, such as Teach for America and Troops to Teachers, do not certify teachers, but rather recruit non-traditional teaching candidates and funnel them into certification programs. Teachers who obtain their certification through either a traditional or an alternative route need to be distinguished from teachers who hold emergency, temporary, or provisional certification because teachers holding these latter types of certifications are not necessarily in a training program or on their way to certification.

There are several critical distinctions between the traditional certification (TC) and alternative certification (AC) routes, including, most notably, the type of candidates who take the routes, the chronological order of critical milestones along the road to full certification, the background of the person who provides the substance of the training, and the emphasis of the training. Perhaps the most controversial aspect of alternative certification is that before taking their first full-time teaching job, alternatively certified teachers take minimal to no education

---

\(^2\)Definitions and certification titles vary from state to state. In this report, “fully certified teachers” refers to those teachers who completed all of a state’s required course work and passed all state licensing examinations.
courses (such as courses in pedagogy, child development, and classroom management) and generally engage in, at most, a few weeks of practice teaching. While the typical AC program will require AC candidates to attend workshops or take university courses during the first year of teaching (and sometimes second and third years of teaching) they are not fully certified for one to three years after first entering the classroom. In contrast, traditionally certified teachers complete a full battery of education courses, participate in an average of 14 weeks of student teaching, and receive their full certification before becoming a full-time teacher.

Recognizing that ED’s primary concern regarding teacher training relates to its impact on students, we limit our review to studies of the impacts of certification programs on the quantifiable measure of student achievement. Several studies address other “outcomes” that may be affected by the method of certification, such as teacher subject matter knowledge, teacher test scores, evaluations of teaching by mentors or principals, and teacher perceptions or attitudes; however, each of these measures has limitations. Collegiate courses and major often serve as proxies for teacher subject matter knowledge, but these indirect measures are limited by wide variation in what constitutes a course or major. Studies that examine the effect of certification using more direct measures of subject matter knowledge, such as the score a teacher receives on a test, reveal little about the subsequent effect on students.3 Less objective measures of teacher effectiveness are sometimes used as outcomes, include ratings of supervisors, teacher self-reports, and independent observations. However, given the underlying incentive issues and the subjective nature of the outcomes, it is difficult to accord these measures much weight.

3 In fact, a separate literature examines the relationship between teacher subject matter knowledge and student achievement (Monk 1994; Monk and King 1994; Goldhaber and Brewer 1997).
Unfortunately, the literature on alternative certification that focuses directly on student achievement is shockingly small. Wilson et al. (2001) cite 14 articles that examine the issues of alternative certification as meeting their requirements for inclusion. Of those 14, only two include analyses that examine the impact of certification on student achievement; the others evaluate teacher preparation by using teacher ratings, surveys, or interviews as outcomes. Out of numerous other studies, we were able to identify four other articles that broadly meet the requirements for inclusion in this review. Upon further examination, it is clear that each of the six articles is flawed by problems that undermine the study; while some have relatively minor problems, the majority of them contain more significant flaws.

The design of a study is critical to the quality of the evidence it generates. Well-implemented randomized experiments are often considered the gold standard against which to measure other study designs. It is the only approach that can ensure that the effects we see from a program are in fact attributable to that program alone and not to other factors affecting schools and students. None of the six studies reviewed in this paper is based on random assignment, though two used a matched comparison approach. In those papers, the authors attempted to match similar teachers with different types of certification and then looked at the differences in student achievement. The matched comparison approach can be used to create comparison groups that are similar with regard to the characteristics used by the researchers to make the match. However, the approach can have low validity if the set of characteristics used for the match is incomplete or if unobservables cannot be accounted for in the match; that is the case in one of the studies discussed below. The other matched comparison study, which uses a reasonable approach to matching teachers, still results in classes that are not necessarily

4 Table 1 provides a brief summary of the studies included in this review.
equivalent, yet it makes no attempt at regression-adjustment to alleviate the remaining differences.

The four other papers included in this review are based on multiple regression analysis in which student achievement is modeled as a function of teacher certification and other characteristics. The goal is to estimate the relationship between certification and student achievement or how achievement varies with differences in teacher certification after “controlling” for other factors that might affect achievement. For example, students who demonstrate a higher level of achievement (relative to other students) in the year before the study are more likely to reach a higher level of achievement in the year after the study regardless of the type of certification earned by their teacher. If a study does not control for earlier achievement, it may inaccurately attribute differences in post-study achievement to the certification of the teacher. Even with such controls, other unobserved factors are likely to affect achievement and, because they cannot be included in the estimation, will influence the relationships of other variables. Even though the six articles included in this review are the extent of the literature relating alternative certification to student achievement, all are flawed and therefore must be interpreted with caution.

A. MATCHED COMPARISON STUDIES

A study by Miller, McKenna, and McKenna (1998) examines an alternative certification training program created by faculty at a southeastern university. In May 1989, the university started an alternative certification program of individualized and intensive study for 70 middle-grade teachers. The study was designed to comply with Georgia’s provisional certification standards. During the summer of 1989, students took between 15 and 25 quarter-hours (depending on initial assessments) to qualify for provisional certification; 67 participants successfully completed the coursework and were in classrooms in the fall.
In their first year as teachers, participants were supported by additional coursework and received a substantial amount of supervision from a university supervisor and a public school mentor. The university supervisor observed and conferred with the teacher eight times during the year, met with the teacher’s mentor to discuss the progress of the teacher-mentor relationship, and taught a biweekly course for all participants that focused on examining common problems, exploring solutions collaboratively, and providing support. After the first year, support was limited to the additional coursework required to earn regular certification and the informal continuation of the mentor relationship.

Three years after program participation, teachers who were in self-contained fifth and sixth grade classrooms were matched with traditionally certified teachers who began their teaching careers in the same year. The match required that both the traditionally and alternatively certified teachers teach the same subject to students in the same grade and in the same school. The result was a matched sample of 18 total classrooms across 9 schools—in each school, the classroom of one program participant was matched to the classroom of one traditionally certified teacher. Using a multivariate analysis of variance, the authors found no significant difference in the total mathematics or total reading scores (Iowa Test of Basic Skills) that could be attributed to the method of training. In other words, the study suggests that teachers from the alternative certification training program created by faculty at the southeastern university were performing as well as teachers trained in traditional certification programs.

The Miller et al. study is the most convincing in the literature, as the matched comparison methodology comes closest to a random assignment design; however, it is deficient in a number of areas. Primarily, the fact that the results are based on nine participants in one program in a single state severely limits the external validity of the findings. Asserting that the students were not grouped by ability, the authors provide no evidence that students were randomly assigned to
teachers within a grade. The authors collected pretest scores for all students and cite a lack of entry-level differences as justification for excluding covariates from their analysis; however, though the distributions of test scores were similar for the two groups of students, any differences across classrooms (and thus the quality of peers) may have gone ignored by the failure to include pretest scores. Furthermore, the classrooms clearly contained other differences; for example, the nine alternatively certified teachers taught 188 students while the nine traditionally certified teachers taught 157 students, resulting in a class size difference of 3.5 students, on average.

The traditionally certified teachers came from a variety of backgrounds, including states with different training mechanisms. Certainly, it is reasonable to compare teachers from a variety of backgrounds; yet, to ensure a clean comparison between traditionally certified teachers and teachers who pass through the alternative certification program, it is also important to verify that all of the teachers in the comparison group meet the requirements for traditional certification in Georgia. In addition, by limiting the sample to teachers who were still teaching after three years and were “accessible to the campus” of the training program, the methodology may have given rise to selection bias associated with the teachers’ location and career decisions. Comparing the two types of teachers in their third year of teaching makes it more difficult to attribute different student outcomes solely to the method of certification, as other aspects of teacher experience (such as the school or schools in which the teacher was working or the additional coursework taken by a teacher) during the intervening period may be correlated with the results. Similarly, if alternatively and traditionally certified teachers were teaching the same subject to students in the same grade and in the same school over a period of time, their interactions may lead to spillovers in teaching methods or strategies that make it difficult to isolate the effects of participation in the alternative certification program.
Laczko-Kerr and Berliner (2002) provide the only other matched comparison study involving traditionally and nontraditionally certified teachers. The authors combine all teachers who do not meet the Arizona state requirements for certification (a bachelor’s degree from an accredited institution, 45 semester hours of education coursework, and passing scores on the Arizona Educator Proficiency Assessment) into one “under-certified” group as the comparison group for traditionally certified teachers. The group includes those who are labeled as emergency (those who hold a bachelor’s degree from an accredited institution and have little or no educational coursework), temporary (a rarely used designation similar to emergency), and provisional (for those with some or considerable teacher training who are missing certain units or requirements that could earn them a standard certificate). Among the “under-certified” teachers in this study are some from an alternative training program, Teach for America (TFA).

After identifying districts with high percentages of under-certified teachers, the study reports that five out of 24 districts agreed to participate in the research; each of the five districts served an inner-city, largely minority population. In 1998–1999 and 1999–2000, those districts hired 293 new teachers whose files contained the necessary demographic and classroom-level student achievement data. Teachers from each group (certified and under-certified) were matched according to grade level taught, highest degree attained, and year of test administration (1998 or 1999); the 109 matched pairs of teachers in third through eighth grades represented 74 percent of the original sample. The authors report that the students of under-certified teachers (those with emergency, temporary, or probationary certification, including Teach for America teachers) performed significantly worse (20 percent) than students of certified teachers in reading, language arts, and mathematics. Stated another way, students in classes taught by a certified teacher received the equivalent of two more months of academic growth. In addition, students of
TFA teachers did not perform any better or worse than students of other under-certified teachers.5

The major methodological problem with the Laczko-Kerr and Berliner study concerns the matching of teachers. When matches within the same school were impossible, the researchers matched teachers between schools within the same school district or between schools in different districts; the authors do not specify how many teachers were matched within school or within district, but they report that 38 percent of teachers were matched with a teacher in another district. To justify the use of their matched data for the analysis, the authors asserted a number of critical assumptions without providing any data to substantiate their claims, such as teachers in the same school taught similar students, teachers in the same district taught similar students, the assignment of teachers to schools and classrooms did not result in a selection effect (that is, that under-certified teachers were not more likely to be found in the worst schools or classrooms), and class size and student ability did not differ across teachers.

Taken as a whole, the methodology of matching is not necessarily a problem. Frequently, the characteristics of the subjects to be matched are used in some combination to ensure that the baseline characteristics of two groups are similar across the relevant dimensions. In the study in question, however, the authors based the pairing of teachers solely on whether they taught in the same school or district (when possible), thus ignoring differences in schools, classrooms, and teachers. If the matching technique were successful, we would observe few if any statistically significant differences in the baseline characteristics of the two groups after matching. However, the limited evidence on match quality provided by the authors contradicts their assertions and

---

5 In the original sample of 293 teachers, 25 percent of the under-certified teachers were TFA teachers (34 out of 134); the authors do not provide the number of TFA teachers in the final sample of 109 under-certified teachers.
justifications for their matching method: using their only measure of achievement (the “outcome” test scores in 1998 and 1999), they find that, across all schools in the study, reading test scores differ significantly and that, across all districts in the study, test scores for several subjects differ significantly. Therefore, matching based solely on the teachers’ school or district yields matched samples that vary substantially in their characteristics. Furthermore, the authors assert that, for matches across districts, the student populations had similar economic bases; while it may be true that two districts have similar socioeconomic characteristics, it is certainly not the case that each school in the district has those characteristics. The authors present no evidence for the comparability of matched teachers along these socioeconomic dimensions.

Some of the matching and other methodological problems could be alleviated if the authors were able to examine the gain in student achievement by controlling for initial student achievement. However, when the only outcome measure is the level of student achievement at one point in time, the authors cannot draw conclusions regarding teacher effectiveness. In addition to failing to control for differences in baseline achievement, the authors do not control for other classroom characteristics, such as class size. The implicit assumption is that all classrooms are the same such that all differences in final achievement are attributable to the training of the teacher. Given that the authors’ estimation strategy is based on these unverifiable and questionable assumptions, we can have little confidence in the study results.

Overall, the aggregation of TFA teachers with those holding only an emergency credential creates a serious specification error and prevents any insight into the effect of the characteristics of the TFA program. Furthermore, given the sample sizes, isolating one subgroup of the “under-certified” population and comparing it to the others is unreasonable. For example, after splitting the sample by year of test, the number of TFA teachers used for the subgroup analysis was 8 in 1998 and 22 in 1999. The small sample of TFA teachers leads to large standard errors, reducing
the likelihood of finding a significant difference between subgroups of the under-certified teachers.

B. REGRESSION ANALYSIS

Using data from the National Educational Longitudinal Study of 1988, Goldhaber and Brewer (2000) examine how students of teachers with probationary certification, emergency certification, private school certification, or no certification in their subject area compare relative to students of teachers who have earned standard certification in their subject area. Unlike the previous study, the Goldhaber and Brewer study examines alternatives to standard certification in terms of the individual student rather than collectively across all students. The focus of the study is 12th grade standardized test scores in mathematics and science for individual students; the sample consists of 3,786 mathematics and 2,524 science students taught by 2,098 mathematics and 1,371 science teachers in a nationwide sample of public schools. The certification variable was created from responses to a survey question that asked, “Which type of math and science teaching certifications do you hold from the state where you teach?”

To isolate the effect of teacher certification with a high degree of accuracy, the authors control for an extensive set of variables that may also affect student achievement, including individual and family background variables, school variables, teacher variables, and class variables. Using a teacher random effects model (to account for multiple students per teacher), the study finds that, in mathematics, students with an uncertified teacher or a teacher with a private school certification score 1.3 points lower (10 percent of the standard deviation) than those taught by a teacher with a standard, probationary, or emergency certificate. On the other hand, students taught by a teacher with an uncertified teacher or a teacher with a private school certification score 1.3 points lower (10 percent of the standard deviation) than those taught by a teacher with a standard, probationary, or emergency certificate.

6 The response categories were regular or standard, probationary, emergency, private school certification, and not certified in subject. Thus, there is no distinction between those who are not certified and those who are not certified in their subject area.
hand, the study finds no evidence that, with respect to student achievement, teachers with a standard certificate in their subject outperform those with an emergency certificate. Teachers who have taken an alternative route to certification may fall in many of these categories; however, if an alternative certificate is considered at least as good as an emergency certificate, the findings lend support to those who advocate that alternative certification is a feasible alternative to the traditional route.

Based on the already noted survey question, the data cannot distinguish between those who are completely uncertified and those who are not certified in their subject area; thus, teachers who are certified but teaching out-of-field are grouped with the uncertified teachers. It is therefore possible that the findings for the uncertified teachers are positively biased if the certified, out-of-field teachers outperform the other uncertified teachers. The data are also limited by variation in the definition of certification across states and by potential measurement error associated with variation in the interpretation of the survey question across individual teachers.

Furthermore, none of the nonstandard certifications (probationary, emergency, private school, or no certification) corresponds strictly to a particular program of alternative certification. The data are incapable of distinguishing the route through which teachers entered the teaching profession, and they cannot describe the qualifications associated with a particular response to the survey question. Therefore, the study cannot directly explore the components of certification that affect student achievement. In addition, given that some of the teachers have undoubtedly completed alternative certification programs, it is possible that they differ in terms of unobservables (such as motivation) that are impossible to quantify in the data. Thus, any differences in student performance may be attributable to either the impact of the training received by the teacher or a teacher selection effect.
Beyond the study’s inability to identify particular alternative certification programs or characteristics associated with student achievement, several issues call into question the strength of the findings. Given that the nonstandard certification categories include those who are certified but teaching out-of-field, the reported percentage of teachers classified as such (less than 7 percent) appears questionably low (Seastrom et al. 2002). Even if accurate, the sample sizes by subject and certification type are still remarkably small; for example, the result from the comparison of emergency to standard certification in mathematics was based on 49 students taught by teachers with emergency certification and 3,179 students taught by teachers with standard certification. Another issue arises from basing the impact of the 12th grade teacher’s certification on 12th grade test scores when 10th grade tests provide the information on earlier achievement. In the intervening period, students are likely to be influenced by multiple teachers across multiple courses. Using information only from the last teacher biases the estimates by attributing the entire difference to a single teacher. Furthermore, if the assignment of a student to a teacher is based at all on student performance, selection bias will occur along this dimension as well.

Raymond, Fletcher, and Luque (2002) seek to examine the effect of Teach for America (TFA) teachers on student achievement in Houston. During the past 12 years, the Teach for America program has provided an alternative supply of teachers in some of the nation’s largest and most needy school systems.7 Candidates for the program, who are recruited from more than 200 colleges and universities, undergo screening by TFA staff. The screen includes writing an essay, participating in a personal interview, and conducting a sample teaching session.

7 TFA communities include Atlanta, Baltimore, the San Francisco Bay Area, Chicago, Houston, Los Angeles, the Mississippi Delta, New Jersey, New Orleans, New York City, North Carolina, Phoenix, the Rio Grande Valley in Texas, rural Louisiana, and Washington, DC.
Candidates who pass the screen attend a five-week summer institute operated by TFA in conjunction with the Houston Independent School District. At the institute, candidates work in teams to teach summer classes under the supervision of experienced teachers – usually TFA alumni. In addition to working directly with students in classrooms, TFA candidates participate in professional development activities that emphasize topics such as managing classrooms, assessing student performance, and motivating students and families to sustain high levels of academic performance.

Using data from the Texas Assessment of Academic Skills, the authors were able to control for earlier student achievement in their study of 186 elementary and 34 middle schools. Estimating a teacher fixed-effects model to generate average gains for each teacher, they attempted eight sets of analyses based on subject (mathematics and reading), grade level (elementary and middle), and comparison group (all non–TFA teachers and new non–TFA teachers). Within the school level, the authors pooled data across grades, asserting that the underlying learning process is stable from year to year. After controlling for school, class, teacher, and student characteristics, they found that, on average, TFA teachers produced a positive effect on their students’ achievement levels, though the differences were generally not statistically significant.

Of the elementary results, the two analyses using all non–TFA teachers as the comparison group were small and positive though insignificant while the two analyses using only the new non–TFA teachers as the comparison group found larger and positive results, with a significant effect in mathematics. In the middle grades, three of the four analyses resulted in positive and significant findings when a dummy variable was used for whether a student had a TFA teacher; only the positive impact on reading test scores for TFA teachers as compared to new non–TFA teachers was not statistically significant. None of the findings proved statistically significant.
when the TFA intensity variable (proportion of time taught by a TFA teacher) was used to account for the possibility that a student had been taught by several teachers with a variety of backgrounds. In general, the distribution of test scores for TFA teachers was narrower, suggesting greater consistency and less variability. Furthermore, the mean of the distribution was always higher than that of the non–TFA teachers.

The elementary findings are based on the extremely broad comparison of TFA teachers with all non–TFA teachers in the Houston school system in the fourth and fifth grades. As demonstrated by the authors, the characteristics of the average school for TFA teachers differ considerably from those for non–TFA teachers; for example, TFA teachers tend to work in higher-poverty schools. In addition, many of Houston’s elementary schools contain departmentalized fourth and fifth grade classrooms, creating the same attribution problem as in the middle school analyses. While the statistical model can control for some of these measurable differences, unmeasured differences are probably correlated as well. Furthermore, principals often systematically steer certain types of students to TFA or non–TFA teachers.8

At the beginning of their first year, new TFA teachers in Houston had to enroll in the district’s large and popular Alternative Certification Program (ACP), through which they were assigned a mentor, attended weekly training sessions, had monthly observations and consultations with an ACP specialist, and were required to take two courses at a local university. However, it is unclear how many of the non–TFA teachers were also enrolled in the ACP program and thus received similar support. If the non–TFA teachers were not enrolled in the

---

8 In the feasibility and design phases of our study of TFA in Houston, Mathematica Policy Research determined that several Houston elementary schools departmentalize their fourth and fifth grade classrooms, and that principals frequently assigned students based on teacher training.
ACP program, some of the gains attributed to TFA could have come instead from the ACP program.

Two other regression studies have attempted to address the relationship between certification and student achievement; however, rather than using student- or teacher-level data, the studies used data aggregated to the school or state level. While we discuss their findings below, it should be noted that aggregation bias might produce significantly distorted results.

Using data from 795 regular California high schools, Fetler (1999) examined the relationship between measures of teacher certification and student achievement. Specifically, he focused on the effects of teachers with an emergency permit (holding a bachelor’s degree, passing a basic skills test, and completing at least 18 semester hours or 9 upper-division/graduate semester units of course work in mathematics), a limited-assignment emergency permit (holding a valid teaching credential in another subject), or a waiver (passing the mathematics portion of a basic skills test). The model estimated the relationship between grade-level mean achievement test scores from the mathematics portion of the Stanford Achievement Test Series and the percent of mathematics teachers with emergency certification in the school. It found that a higher percentage of emergency certifications was associated with lower test scores; however, no tests were performed to assess whether the differences were statistically significant.

The author of the study points out the most problematic flaw in the analysis: students at public secondary schools with a higher poverty level or with a higher- percentage minority enrollment were more likely to receive mathematics instruction from a teacher who had not majored in mathematics. Though higher percentages of emergency certificates were associated with lower scores, we cannot conclude that they were the cause. Given the evidence on teacher placement, it may be more likely that the relationship goes the other way--schools with a lower average test score are assigned more teachers with emergency certificates.
The disparate levels of aggregation (grade for test scores and school for teachers) are problematic because they provide no direct link between student achievement and teacher training. For example, a school may have two grades in which teachers with standard certificates in mathematics teach mathematics classes and two grades in which teachers with emergency certificates teach mathematics classes. In the analysis in question, all four grades, and thus their grade-level mathematics scores, would be associated with the same level of emergency teaching—that of the entire school. Furthermore, the measure of emergency certification is rough; it is calculated as the number of teachers with an emergency certificate who teach mathematics as a share of all teachers who teach one or more mathematics classes. Given this specification, teachers receive equal weight independent of the number of mathematics courses taught, the types of mathematics courses taught, the number of students taught, and the grade level of students taught. Again, the lack of control for initial achievement means that it is impossible to attribute anything to the teacher characteristics at one point in time as baseline differences across both classes and schools probably exist. Furthermore, characteristics of the students’ entire course of study affect their test scores, not just the composition of the teaching staff at the time of test administration.

In a paper that has generated much discussion, Darling-Hammond (2000) used data from the 1993–1994 Schools and Staffing Survey and several administrations of the National Assessment of Educational Progress (NAEP) (1990, 1992, and 1996) to examine the effect of certification status on student achievement at the state level. The study examined mathematics scores (fourth grade in 1990 and 1996 and eighth grade in 1992 and 1996) and reading scores (fourth grade in 1992 and 1994) in the public schools of the 44 states that participated in the state NAEP. For each of the six analyses, Darling-Hammond examined the relationship between the average NAEP score for the state and the percentage of well-qualified teachers (those with state
certification and a major in their field), the percentage of teachers with a master’s degree, and the percentage of unqualified newly hired teachers (those with no certificate and those with provisional, temporary, or emergency certification) while controlling for state average class size and the state percentages of students with incomes below the poverty line and with limited English proficiency.\(^9\) Across the six analyses, Darling-Hammond found that the percentage of well-qualified teachers in the state had a positive and significant effect on state average achievement scores.

Again, as in the Laczko-Kerr and Berliner (2002) and Fetler (1999) studies, the lack of control for earlier achievement is a serious limitation. It is impossible to isolate the effect of teachers on student achievement by examining only one point in time, ignoring all previous inputs into the students’ level of knowledge. In addition to the variation in student ability and observable characteristics within a state, many other unmeasured variables will likely explain why scores vary across states. For example, if the wealthiest states report higher test scores and are more likely to employ well-trained teachers, the two measures of interest will be correlated, both affected by wealth.

Furthermore, by using state-level data, the study exacerbated the bias attributable to aggregation. For example, the use of an average score to represent all student outcomes in a state implies that all students in the state are similarly affected by the characteristics of teachers at the state level rather than by the characteristics of teachers in their district, school, or classroom; failure to account for these variables undermines the study’s implications that there is a link between teacher certification and student achievement. Even if we were to ignore these

\(^9\) Thus, the comparison group was comprised of teachers who were lacking standard certification, a major in field, or both, as long as they were not a new teacher with either no certificate or a provisional, temporary, or emergency certificate.
issues, the findings attribute too much of the difference between state test scores to certification. In the study, a well-qualified teacher is defined as one with state certification and a major in the field. Just as other studies have found a relationship between subject knowledge and student achievement (Monk 1994; Monk and King 1994; Goldhaber and Brewer 1997), it is impossible to disentangle the effects of the two measures of teacher background.

C. CONCLUSIONS

Only a handful of studies examine the relationship between teacher certification and student outcomes. Of those, just two look directly at alternative certification while the others examine the effects of a teacher with standard certification relative to teachers without standard certification, such as those who are uncertified or have an emergency certification. The two alternative certification studies tell us that (1) there was no difference in test scores that could be attributed to a small alternative certification program in Georgia and (2) Teach for America has a positive, though generally insignificant, effect on student achievement in Houston relative to other new non-TFA teachers. The other studies tell us that (1) individual 12th grade mathematics achievement falls with an uncertified teacher despite no difference between teachers with standard and emergency certification, (2) traditionally certified teachers in Arizona elementary and middle schools raise student achievement by 20 percent compared to uncertified teachers, (3) a higher share of emergency certified mathematics teachers in California high schools is associated with lower mathematics scores, and (4) states with higher percentages of teachers with both a standard certification and a degree in field are associated with higher fourth and eighth grade state-average test scores.

However, the findings from all six of these studies are suspect, due to methodological flaws that cannot be overlooked. Both the alternative certification studies and the more generic certification studies use a variety of designs and analytic techniques that yield questionable
findings regarding the effect of certification programs on student achievement. Selection bias, or unmeasurable differences in students across different types of teachers, is a potential problem in almost every study reviewed. Although two studies attempted to address the bias issue by using matched comparison groups, one was left with a small sample on which to base its results, and the other appeared to have different populations after matching. Furthermore, if students with different teachers vary in ways that cannot be measured, the matching studies may still provide unreliable estimates.

As an alternative to the matched comparison design, several studies use multivariate regression specifications to examine the relationship between teacher certification and student achievement. However, each study has difficulty isolating the effect of the teacher’s certification status on student achievement. While some studies are able to control for earlier achievement at the level of the individual student, they are unable to control for students’ exposure to a number of teachers and courses before the final test. Other studies aggregate to the levels above the student or classroom or use coarse measures of teacher certification.

As a whole, the existing literature suggests the need for a more rigorous study in order to assess with a high degree of accuracy the effectiveness of alternative certification programs on students’ achievement. A major component of such a study would be the testing of students before and after exposure to a teacher. Given the infeasibility of randomly assigning teachers to a particular training program or path to certification, a reasonable approach would be to use random assignment in the placement of students in classes taught by teachers with different backgrounds. To evaluate programs of alternative certification with greater precision, a study would need to focus on a few clearly defined alternatives, with detailed components and requirements. However, the study would also need to be large enough to detect reasonably sized impacts and broad enough to provide insight into implications for educational policy.
<table>
<thead>
<tr>
<th>Study</th>
<th>Type</th>
<th>Sample</th>
<th>Comparison Group</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Darling-Hammond (2000)</td>
<td>Regression</td>
<td>National (NAEP): 44 states 4th and 8th grade</td>
<td>Unqualified newly hired teachers (those with no certificate and those with provisional, temporary, or emergency certification) versus well-qualified (those with state certification and a major in their field)</td>
<td>The percent of teachers in a state who were well-qualified had a positive and significant effect on state average achievement test scores.</td>
</tr>
<tr>
<td>Fetler (1999)</td>
<td>Regression</td>
<td>California: 795 high schools</td>
<td>Share of teachers in a school with emergency permits</td>
<td>A higher percent of emergency certifications in a school was associated with lower test scores.</td>
</tr>
<tr>
<td>Goldhaber and Brewer (2000)</td>
<td>Regression</td>
<td>National (NELS) :</td>
<td>Separate analyses for probationary certification, emergency certification, private school certification, or no certification versus standard certification</td>
<td>In math, students with an uncertified teacher or a teacher with a private school certification score 1.3 points lower on standardized test (relative to standard, probationary, or emergency). No evidence that teachers with a standard certificate in their subject outperform those with an emergency certificate.</td>
</tr>
<tr>
<td>Laczko-Kerr and Berliner (2002)</td>
<td>Matched Comparison</td>
<td>Arizona: 109 matched pairs 3rd through 8th grades</td>
<td>Under-certified (all who did not meet state requirements for certification - emergency, temporary, probationary, and TFA) versus certified</td>
<td>Under-certified did 20 percent worse on reading, language arts, and math. Students of TFA teachers performed no better or worse than the other under-certified teachers.</td>
</tr>
<tr>
<td>Miller, McKenna, and McKenna (1998)</td>
<td>Matched Comparison</td>
<td>Georgia: 9 matched pairs 5th and 6th grades</td>
<td>Alternative Certification Program versus traditional certification</td>
<td>No difference in total math or total reading scores.</td>
</tr>
<tr>
<td>Raymond, Fletcher, and Luque (2002)</td>
<td>Regression</td>
<td>Houston, Texas: 186 elementary &amp; 34 middle schools</td>
<td>Teach For America versus non-TFA (all and newly hired)</td>
<td>TFA teachers produced a positive effect on their students' achievement levels, though the differences were generally not statistically significant. The distribution of test scores for TFA teachers had both a higher mean and lower variance, suggesting more consistent test scores.</td>
</tr>
</tbody>
</table>
REFERENCES


Seastrom, Marilyn McMillen, Kerry J. Gruber, Robin Henke, Daniel J. McGrath, and Benjamin A. Cohen. *Qualifications of the Public School Teacher Workforce: Prevalence of Out-of-