

Defining “Highly Qualified Teachers”: What Does “Scientifically-Based Research” Actually Tell Us?

by Linda Darling-Hammond and Peter Youngs

In July 2002, the U.S. Secretary of Education issued the Secretary’s Annual Report on Teacher Quality (U.S. Department of Education) as required by the 1998 reauthorization of Title II of the Higher Education Act. In this report titled *Meeting the Highly Qualified Teachers Challenge* the Secretary essentially argues for the dismantling of teacher education systems and the redefinition of teacher qualifications to include little preparation for teaching. Stating that current teacher certification systems are “broken,” and that they impose “burdensome requirements” for education coursework that make up “the bulk of current teacher certification regimes” (p. 8), the report argues that certification should be redefined to emphasize higher standards for verbal ability and content knowledge and to de-emphasize requirements for education coursework—making student teaching and attendance at schools of education optional and eliminating “other bureaucratic hurdles” (p. 19). These conclusions rest on the following arguments, each of which is addressed in turn in this article:

- Teachers matter for student achievement, but teacher education and certification are not related to teacher effectiveness.
- Verbal ability and subject matter knowledge are the most important components of teacher effectiveness.
- Teachers who have completed teacher education programs are academically

weak and are underprepared for their jobs.

- Alternative certification programs (ACPs) have academically stronger recruits who are highly effective and have high rates of teacher retention.

The report suggests that its recommendations are based on “solid research.” However, none of these arguments has strong empirical support, and the report does not cite the scientific literature that addresses them: Only one reference among the report’s 44 footnotes is to a study that was eventually published in a peer-reviewed journal, and the study’s findings are misrepresented in the report. Most references are to newspaper articles or to documents published by advocacy organizations, some of these known for their vigorous opposition to teacher education.¹

Although an accurate review of rigorous research on teacher qualifications and their relationship to student achievement could provide useful guidance to state policymakers, such a review is not to be found in this report. Instead, the Secretary’s report fails to meet the Department of Education’s own standards for the use of scientifically based research to formulate policy. The report cites almost no research that would meet scientific standards, misrepresents findings from a large number of sources, and includes many unsupported statements about teacher education and teacher certification. Whatever the contributions of this report to the debates on teacher quality, an accurate rendering of the research base on these important topics is not one of them. In this article we discuss the research base that treats the arguments made in support of the report’s recommendations and suggest that different conclusions would derive from a well-grounded rendering of the evidence.

Proposition I: Teachers Matter for Student Achievement, but Teacher Education and Certification Are not Related to Teacher Effectiveness

The Secretary’s report accurately claims that “researchers have found that some teachers are much more effective than others” (2002, p. 7). Studies using value-added student achievement data have found that student achievement gains are much more influenced by a student’s assigned teacher than other factors like class size and class composition (Sanders & Horn, 1994; Sanders & Rivers, 1996; Wright, Horn, & Sanders, 1997). A recent analysis by Rivkin, Hanushek, and Kain (2001) attributes at least 7% of the total variance in test-score gains to differences in teachers.

The Secretary’s report asserts, however, that “there is little evidence that education school course work leads to improved student achievement” (2002, p. 19), stating that the evidence about “knowledge of pedagogy, degrees in education or amount of time spent practice teaching”—which are the “requirements that make up the bulk of current teacher certification regimes”—is surrounded by a “great deal of contention” (p. 8). To support the assertion that “virtually all” of the studies linking certification and improved student outcomes are “not scientifically rigorous,” the Secretary’s report cites a report by Kate Walsh (2001), written for the Baltimore-based Abell Foundation,² which asserts that there is “no credible research that supports the use of teacher certification as a regulatory barrier to teaching” (p. 5). Unfortunately, Walsh’s report excludes much of the evidence on the topic, misrepresents many research findings, makes inaccurate claims about studies that have examined the consequences

The Research News and Comment section publishes commentary and analyses on trends, policies, utilization, and controversies in educational research. Like the articles and reviews in the Features and Book Review sections of *ER*, this material does not necessarily reflect the views of AERA nor is it endorsed by the organization.

of preparation, and uses a double standard in evaluating the research (see Darling-Hammond, 2001, 2002).

Walsh's report was written both to present a case against Maryland's efforts to strengthen teacher preparation requirements and to defend the continuation of a local short-term alternative route into teaching that had come under criticism. It also attacks the state's efforts to require courses in the teaching of reading for all teachers, characterizing them as additional "barriers" to the ability to teach. Suggesting that state and local policymakers are misguided in their efforts to seek more fully certified teachers for Baltimore City schools, Walsh's report ridicules the reports of local journalists and a Baltimore community group that released data that "bemoaned the fact that more uncertified teachers were teaching in the city's high-poverty, predominantly African-American schools than the city's whiter, more affluent schools" (p. 2). The report contends that these inequalities in access to certified teachers are not problematic if certification is discounted as a determinant of achievement.

Defining Scientifically Based Research

The Secretary's report dismisses the importance of teacher preparation by arguing that the research linking teacher preparation to measures of teacher effectiveness is scientifically inadequate, referencing Walsh's report for this assertion. Walsh seeks to marginalize much of the scientific research on teacher education by suggesting it is inappropriate to cite studies that are older, have relatively small samples, use measures of performance other than student achievement scores, are aggregated at a level above the individual teacher or classroom, or have been published in venues other than peer-reviewed journals. Although Walsh rejects or ignores research findings that suggest the influence of teacher education on student learning, she cites comparable research—sometimes the very same studies—when they agree with her assertions about verbal ability or content knowledge (e.g., Ferguson, 1991; Fetler, 1999; Goldhaber & Brewer, 2000; Hawk, Coble, & Swanson, 1985; Monk, 1994; Strauss & Sawyer, 1986). She also cites studies that do not meet the criteria for age, size, outcome measures, aggregation level, or publication venue she uses to discredit studies whose findings challenge her con-

clusions. Ultimately, she is unable to provide well-grounded criticisms of a number of methodologically strong studies (some of them reviewed here), which show strong relationships between student achievement and teachers' professional preparation and certification status.

Indeed, a recent review commissioned by the Office of Educational Research and Improvement that was vetted for scientific rigor by a panel of researchers disagrees with both Walsh's conclusions and the Secretary's report. This review, which analyzes 57 studies that met specific research criteria and were published after 1980 in peer-reviewed journals, concludes that the available evidence demonstrates a relationship between teacher education and teacher effectiveness (Wilson, Floden, & Ferrini-Mundy, 2001). The review documents relationships between teacher qualifications and student achievement across studies using different units of analysis and different measures of preparation and in studies that employ controls for students' socioeconomic status and prior academic performance.

To evaluate the weight of evidence in a field it is often necessary to triangulate findings that used different methods, over different time periods, and at different levels of aggregation to see how evidence has accrued over time and across methods. Of course, it is important to do this with attention to the methodological strengths and weaknesses of various studies and lines of research. We address some of the methodological issues that have surfaced as arguments for discounting the results of the body of research on teacher preparation.

Age of Studies. Walsh's report discounts a large number of studies of teacher education and certification because they were conducted before 1980. Yet, this is when most studies using experimental or matched comparison designs were completed. This is in part because there was great variability in entry pathways and much interest in the topic during the high-demand years of the 1960s and 1970s and because federal funding for educational research was substantially larger before 1980 than it has been since. Fewer studies were concerned with the issues of preparation and certification in the 1980s when virtually all teachers were certified, and few data sets included measures of teacher education variables. Inter-

est and data on this topic just began to return in the 1990s. Although some newer data sets provide more useful information on questions of teacher education and certification, federal education data sets are updated relatively infrequently and are made available even more slowly. Although the age of studies can sometimes influence their applicability to current contexts, and they must be interpreted with these questions in mind, they do not become invalid merely because they are old. Walsh cites numerous studies that are very old—for example, Bowles and Levin (1968), Coleman et al. (1966), and Massey and Vineyard (1958) for the proposition that verbal ability matters (Walsh, 2001, p. 6)—even though she dismisses some of them in her separately published appendix as "too old."

Sample Size and Methods. Another argument used to discount many studies is the size of their samples. Walsh bemoans the lack of experimental research, but she then rejects the results of studies with experimental and matched comparison designs because of their relatively small sample sizes. This is true except when she agrees with the findings of particular studies. For example, although Walsh summarily dismisses a set of studies with sample sizes of 40 or fewer teachers (p. 25), she cites Miller, McKenna, and McKenna (1998), a study that included student achievement data from only 18 teachers, for her proposition that "new teachers who are certified do not produce greater student gains than new teachers who are not certified" (p. 8). (In fact, as we will discuss, this study did not include uncertified teachers, but was a study of an alternative teacher education program.) In the original version of her report,³ she also cites Bullough, Knowles, and Crow (1989), a study of three student teachers, for her proposition that preservice training "fades quickly from new teachers' minds" and Hawk, Coble, and Swanson's (1985) matched comparison study of 36 teachers as evidence for the influence of teachers' subject matter knowledge on student achievement (although she discounts its findings for the effects of certification). Her arguments that certification rules should be relaxed are made on the basis of anecdotes about three individual teachers (pp. 38–40).

The larger correlational studies on which Walsh often relies typically do not include variables that measure teacher education

directly, lack direct controls, and must rely on statistical manipulations of data to account (indirectly) for these other influences. This kind of correlational research is, of course, legitimate for staking out broad possibilities in relationships among variables, but it has its own limitations. Many experimental designs and matched comparison studies can in fact offer more solid evidence about effects because the “treatment” they are studying is known and the samples can be better controlled than in studies that use proxies and statistical controls rather than direct observation of the phenomena of interest.

Medical research, for example, typically uses small sample experimental research as the basis for establishing the possibilities of effects and uses large correlational studies as rough indicators of possible relationships that require further examination. Single case studies of clinical findings are part of the medical research base along with small, carefully controlled experiments, small and large clinical trials, and correlational studies looking at broad tendencies. Although medical researchers generally consider correlational studies to compose a weaker source of definitive evidence about effects than experimental designs, researchers recognize that mixed methods of research serve complementary purposes. For example, the January 2002 issue of the prestigious *New England Journal of Medicine* includes a study with a sample of eight patients who received cardiac transplants (Quaini et al., 2002) and a study of 53 children infected with *E Coli* (Chandler et al., 2002)—neither of which had experimental designs—along with a study of 750,000 Norwegian women whose birth outcomes were examined via medical records (Skjoerven, Wilcox, & Lie, 2002). The usefulness of small, comparison group studies—as well as large correlational studies that use grosser measures—is not in the definitiveness of their individual findings but in their contribution to a larger body of work from which evidence can be triangulated.

Of course, one of the reasons correlational studies must be interpreted with caution is that there is always the question of which direction the correlations may point, sometimes referred to as “reverse causation.” There is also the problem that variables in these studies are frequently crude proxies for the actual measures of interest and may either fail to capture the in-

tended construct or in fact be reflecting the influences of other unmeasured variables. For example, many studies finding strong influences of measures of teacher verbal ability on student achievement have lacked other measures of teachers’ preparation that, when examined in other studies, are also strong predictors. Furthermore, many of the variables that reflect teacher quality are highly correlated with one another—for example, teachers’ education levels are typically correlated with age, experience, and general academic ability, and certification status is often correlated with content background as well as education training and experience (e.g., Goldhaber & Brewer, 2000).

The effects reflected by any given variable in a particular study depend on whether other variables that may also measure aspects of competence are represented in the estimates. The effect size also depends on other context factors, such as the range of variability in the measure used, which can change in different locations and time periods. For example, in some eras and in some locations virtually all teachers held content degrees or were fully certified, so these variables do not strongly predict variations in outcomes. When much more variability is present, these variables are strongly predictive of outcomes. Thus, several studies have found strong measured influences of certification status on student achievement in states like California and Texas during the 1990s when there were wide differences in teachers’ qualifications. For all of these reasons, it is critical for any review of research to represent a range of studies that can shed light on the different relationships of interest using a variety of measures.

Level of Aggregation. Another criticism used to dismiss some studies’ findings as irrelevant is the charge of “aggregation bias.” For example, Walsh dismisses studies that include favorable findings about the value of teacher education in which data are aggregated at the level of the school or district, although she cites similarly aggregated data for her conclusion that verbal ability matters most (e.g., Coleman et al., 1966; Ferguson, 1991; Strauss & Sawyer, 1986). More important, this critique misses a crucial point about how research results accrue and are triangulated to look at possible relationships among conditions and outcomes. Just as individual-level data about health

practices and outcomes inform medical research, so do highly aggregated data at the level of cities, counties, and even countries when researchers seek to understand, for example, why women in some nations have low levels of breast cancer or men have low levels of heart disease.

Although the size of measured effects of different variables can vary at different levels of the system, it is not always clear how the bias operates. Often, the general direction of the results holds at different levels of the system, even if effect sizes differ. For example, Ferguson and Ladd (1996) found the effects on student achievement of teachers’ test scores, master’s degrees, and experience held at both the district and school levels in terms of significance and directionality. There are advantages and limitations for different levels of analyses. On the one hand, disaggregated data can exhibit greater measurement error. On the other hand, some analysts have argued that omitted variables may bias the coefficients of school input variables upward when data are aggregated to the district or state level (Hanushek, Rivkin, & Taylor, 1995). However, this generalization does not always prove true. For example, Summers and Wolfe (1975) found that selectivity ratings of each teacher’s undergraduate institution were important in explaining sixth-grade students’ achievement when examined at the individual teacher level; however, this relationship disappeared when the authors aggregated the college ratings and other school inputs into school-level averages. This contradicts the assumption about the usual direction of aggregation bias.

Of course, omitted variables can bias results at any level of the system. Sometimes, especially when the goal of a study is to evaluate broad trends and policy influences, it is important to have data aggregated and analyzed at multiple levels. For interpreting the weight of evidence on a particular issue, the most important question is whether consistent results are found at different levels of aggregation. With these concerns in mind, we discuss the actual findings of research on the questions raised in the Secretary’s report.

Research on Teacher Education and Certification

A variety of teacher experiences and attributes appear to contribute to the effects that teachers have on student learning.

Looking across studies, several aspects of teachers' qualifications have been found to bear some relationship to student achievement. These include teachers' (a) general academic and verbal ability; (b) subject matter knowledge; (c) knowledge about teaching and learning as reflected in teacher education courses or preparation experiences; (d) teaching experience; and (e) the combined set of qualifications measured by teacher certification, which includes most of the preceding factors (Darling-Hammond, 2000).

As the state's legal vehicle for establishing competence for members of professions, including teaching, licensing, or certification is meant to represent the minimum standard for responsible practice. Current requirements for licensing include measures of many of the variables we noted, such as basic skills and general academic ability, knowledge about subject matter, knowledge about teaching and learning, and some teaching experience.⁴ Over the past decade, states have taken steps to strengthen their licensure requirements, which are now substantially stronger than they were 15 years ago. In most states, candidates for teaching must earn a minimum grade point average or achieve a minimum test score on tests of basic skills, or general academic ability or general knowledge (or both) in order to be admitted to teacher education or gain a credential. In addition, they must generally secure a major or minor in the subject(s) to be taught or pass a subject matter test, take specified courses in education, and, in some states, pass a test of teaching knowledge and skill. In the course of teacher education and student teaching, candidates are typically judged on their teaching skill, professional conduct, and the appropriateness of their interactions with children.

Studies employing national, state, and other data sets have reported significant relationships between teacher education and certification measures and student performance at the levels of the individual teacher (Goldhaber & Brewer, 2000; Hawk, Coble, & Swanson, 1985; Monk, 1994); the school (Betts, Rueben, & Dannenberg, 2000; Fetler, 1999; Fuller, 1998, 2000); the school district (Ferguson, 1991; Strauss & Sawyer, 1986); and the state (Darling-Hammond, 2000). The convergence of findings in studies using different

units of analysis reinforces the strength of inferences that might be drawn from any single study.

Individual Teacher-Level Data. The only study cited in the Secretary's report that was eventually published in a peer-reviewed journal is Goldhaber and Brewer's (2000) examination of the relationship between teacher qualifications and student achievement using data from the National Educational Longitudinal Studies (NELS) of 1988.⁵ The Secretary's report cites this study as its only reference for an inaccurate statement that subject matter degrees have a greater effect on teacher effectiveness than certification:

Research has generally shown that high school math and science teachers who have a major in the subjects they teach elicit greater gains from their students than out-of-field teachers, controlling for student's [*sic*] prior academic achievement and socioeconomic status. These same studies also suggest that possessing an undergraduate major in math and science has a greater positive effect on student performance than certification in those subjects. (p. 8)

In fact, Goldhaber and Brewer (2000) found strong influences of teacher certification on student achievement in high school mathematics and science, above and beyond the effects of teachers' subject matter degrees. They report:

We find that the type (standard, emergency, etc.) of certification a teacher holds is an important determinant of student outcomes. In mathematics, we find that students of teachers who are either not certified in their subject (in these data we cannot distinguish between no certification and certification out of subject area) or hold a private school certification do less well than students whose teachers hold a standard, probationary, or emergency certification in math. Roughly speaking, having a teacher with a standard certification in mathematics rather than a private school certification or a certification out of subject results in at least a 1.3 point increase in the mathematics test. This is equivalent to about 10% of the standard deviation on the 12th-grade test, *a little more than the impact of having a teacher with a BA and MA in mathematics* [*italics added*]. Though the effects are not as strong in magnitude or statistical significance, the pattern of results in science mimics that in mathematics. Teachers who hold private

school certification or are not certified in their subject area have a negative (though not statistically significant) impact on science test scores. (p. 139)

The effect of certified teachers on student achievement was larger in both mathematics and science than the effect of content degrees at the bachelor's and master's degree levels. The fact that the study found a large effect of certification status after controlling for content major suggests that what certified teachers learn about teaching adds to what they gain from a strong subject matter background.

The Secretary's report misinterprets the findings from this study yet a second time stating, "there was no statistical difference in performance between teachers who attended conventional training programs and received traditional teaching licenses versus those who did not complete such programs and were teaching on emergency or temporary certificates" (2002, p. 8). In fact, Goldhaber and Brewer's study does not include data about which teachers had attended "conventional" or other training programs or which had received "traditional teaching licenses." NELS only included information on the type of certificate teachers held in the specific mathematics or science field taught.

The study did find that students of the sample's small number of science teachers with temporary or emergency certification in the science field they were teaching (24 out of the 3,469 teachers in the overall sample) did no worse than the students of teachers holding standard certification in that field; however, both groups of students did better than the students of uncertified teachers. Another analysis of these data (Darling-Hammond, Berry, & Thoreson, 2001) shows that in this sample, about two thirds of the teachers on temporary or emergency certificates were experienced and had education training comparable to that of the certified teachers, suggesting that they had likely completed teacher education programs. The pattern of their qualifications and experiences suggested many were already licensed teachers from out-of-state who were typically hired on a temporary license while they secured a new state license, and some were experienced teachers teaching out of their main field, which was frequently another mathematics or science field.⁶ Only a third of this sample were new teachers whose characteristics

suggested they had a content background with little education training, as the Secretary's report presumes. In an analysis of covariance that controlled for students' pretest scores and teachers' degrees and experience, the students of this subsample of teachers had lower achievement than those of the more experienced and traditionally trained teachers (Darling-Hammond, Berry, & Thoreson, 2001).

Other research on teacher certification at the individual teacher-level is consistent with these findings. In a matched comparison group study of 36 middle school mathematics teachers and 826 students in North Carolina where teachers were matched by years of experience and school setting, Hawk, Coble, and Swanson (1985) found that the students of fully certified mathematics teachers experienced significantly higher gains in achievement than those taught by teachers not certified in mathematics. The differences in student gains were greater for algebra classes than general mathematics.

Teachers' education coursework has also been found to add to the influences of subject matter knowledge in predicting student achievement. For example, using data on more than 2,800 students from the Longitudinal Study of American Youth (LSAY), Monk (1994) found that teachers' college coursework in the subject field was usually positively related to student achievement in mathematics and science, and education courses in subject matter methods had a positive effect on student learning at each grade level in both fields. In mathematics, these methods courses had "more powerful effects than additional preparation in the content area" (p. 142). Monk concludes that, "a good grasp of one's subject area is a necessary but not a sufficient condition for effective teaching" (p. 142).

More recently, Wenglinsky (2000) used data from the National Assessment of Educational Progress (NAEP) to examine the relationships between teachers' training, teaching practices, and student achievement, controlling for student characteristics and other school inputs. He found that eighth-grade students do better on the NAEP mathematics assessments when they have had teachers who engage in more hands-on learning emphasizing higher order thinking and who have a major or minor in mathematics or mathematics ed-

ucation; more professional training in how to work with diverse student populations (a combined measure of training in cultural diversity, teaching limited English proficient students, and teaching students with special needs); and more training in how to develop higher order thinking skills. Similarly, students whose teachers majored in science or science education and had more training in how to develop laboratory skills and engage in more hands-on learning do better on the NAEP science assessments. (The NAEP of 1998 asked teachers to report either college coursework or in-service training in these areas.)

School-Level Data. Several school-level analyses provide further evidence that teachers' certification status is related to student achievement. Three recent school-level studies in California found significant negative relationships between average student scores on the state examinations and the percentage of teachers on emergency permits, after controlling for student socioeconomic status and other school characteristics (Betts, Rueben, & Dannenberg, 2000; Fetler, 1999; Goe, 2002). All of these studies also found smaller positive relationships between student scores and teacher experience levels, with negative effects on student achievement associated with the proportion of beginning teachers. These studies join a number of others in finding that, among school resources, teacher qualifications often appear to have the greatest influence on what students learn and that qualified teachers are unequally allocated to students by race, income, and location.

Similarly, Fuller (1998, 2000) found that students in Texas schools with greater proportions of certified teachers were significantly more likely to pass the Texas Assessment of Academic Skills (TAAS), after controlling for students' socioeconomic status and teacher experience. In one set of studies, he found that the likelihood of elementary school students passing all subtests of the TAAS was greater in schools with higher proportions of certified teachers, controlling for teacher experience, and that gains in pass rates were related to the proportion of properly certified teachers, with prior achievement and student demographics taken into account. The differences were significant for Hispanic students and lower income students (Fuller, 1998).

In a second set of studies, Fuller (2000) found that the percentage of properly certified Algebra I teachers in a school was positively and significantly associated with gains in student achievement after controlling for student and school characteristics.

District-Level Data. Researchers using data aggregated at the district level also report significant relationships between teachers' scores on certification tests and student performance. In a study of nearly 900 Texas school districts that controlled for student background and district characteristics, Ferguson (1991) reports that combined measures of teachers' expertise—scores on a state licensing examination, master's degrees, and experience—accounted for more of the interdistrict variation on students' reading achievement and achievement gains in Grades 1 through 11 than students' race and socioeconomic status. Of the teacher qualification variables, the strongest relationship was found for scores on the state licensing examination, the Texas Examination of Current Administrators and Teachers, which is described by the test developer as measuring basic communication skills, research skills, and teaching knowledge. Master's degrees also exerted a small but significant influence on student achievement, followed by experience.

In another district-level study, Strauss and Sawyer (1986) report that student performance in North Carolina districts was strongly associated with teachers' average scores on the National Teacher Examinations (NTE). The NTE Core Battery featured components measuring teacher's basic skills, general knowledge, and professional teaching knowledge. When the authors controlled for student, school, and community characteristics, they found that teachers' NTE scores had a significant and large effect on students' performance on the state competency examinations in reading and mathematics. In particular, the authors report that a 1% increase in the district average NTE score was associated with a 3% to 5% decline in the district failure rate on the competency exams. They conclude:

Of the inputs which are potentially policy-controllable, . . . our analysis indicates quite clearly that improving the quality of teachers in the classroom will do more for students who are most educationally at risk, those prone to fail, than reducing the class size or improving the capital stock by

any reasonable margin which could be available to policy makers. (1986, p. 47)

Although Walsh (2001) cites this study at least once and Ferguson's (1991) study no fewer than four times in support of her own propositions (pp. 5–7), she dismisses their findings regarding certification by discounting them for “aggregation bias” (p. 27).

State-Level Data. For a study employing state-level data, Darling-Hammond (2000) examined the relative contributions of teacher qualifications, other school inputs, and student characteristics to student achievement across states on the reading and mathematics assessments administered by NAEP in 1990, 1992, 1994, and 1996. After controlling for student poverty and student language background, this study found that measures of teacher preparation and certification were the strongest correlates of average student achievement in reading and mathematics. The most strongly significant predictor of achievement was the proportion of well-qualified teachers, defined as the proportion holding both full certification and a major in the field being taught. The proportion of teachers holding certification exerted an additional small positive effect on achievement and the proportion on emergency credentials exerted an additional small negative effect. The study concludes:

The strength of the “well-qualified teacher” variable may be partly due to the fact that it is a proxy for both strong disciplinary knowledge (a major in the field taught) and substantial knowledge of education (full certification). If the two kinds of knowledge are interdependent as suggested in much of the literature, it makes sense that this variable would be more powerful than either subject matter knowledge or teaching knowledge alone.

In sum, empirical studies employing different units of analysis that have examined the influence of teacher education and certification on student achievement have often found significant relationships between these measures of teacher expertise and student achievement.

Proposition 2: Verbal Ability and Subject Matter Knowledge Are the Most Important Components of Teacher Effectiveness

The Secretary's report asserts, “Rigorous research indicates that verbal ability and

content knowledge are the most important attributes of highly qualified teachers” (2002, p. 19). Although there is research that finds relationships between student achievement and some measures of verbal ability and content knowledge, there is no evidence that these areas of knowledge are more consequential to student achievement than knowledge of teaching. First, most of the studies that have included measures of verbal ability or content knowledge have not included measures of teacher education or certification. Second, in many cases, the relative effect sizes of these measures are no larger than those of teacher education and certification measures.

Evidence of the Importance of Verbal or General Academic Ability

The Secretary's report appropriately claims that “studies have consistently documented the important connection between a teacher's verbal and cognitive abilities and student achievement” (2002, p. 7) but fails to note some important attributes of these studies. The research literature on teacher characteristics has been substantially influenced by the measures available in data sets during particular time periods. Many studies have evaluated the effects of teachers' verbal or general academic ability because these variables have been available in large data sets since the 1960s. On the other hand, data on teachers' content preparation or teacher education experiences have been included in large data sets only since the early 1990s. In a recent review, Wayne and Youngs (in press) found five studies that observed relationships between teachers' verbal or general academic ability and student achievement that met the standard of having controlled for students' socioeconomic status and prior achievement. Four of these studies employed data sets from the 1960s and 1970s and none includes measures of teacher education or certification.

These studies point out how findings with respect to the importance of a particular measure of teacher ability are sensitive to the specification of regressions, as the influences of verbal ability measures trade off with other variables often used as proxies for teacher quality or expertise—college selectivity, other academic ability test scores, higher degrees, and experience.

For example, a study by Hanushek (1992) employed a data set from the Gary Income Maintenance Experiment, a wel-

fare reform experiment in Gary, Indiana in the 1970s. The data set featured reading and vocabulary achievement data on several hundred Black students, most of who were from low-income families, as well as measures of teachers' verbal ability and experience. Controlling for student background characteristics, Hanushek found that teachers' verbal ability scores affected students' reading score gains but not their vocabulary score gains (Hanushek, 1992), which were more strongly influenced by teacher experience. A study by Murnane and Phillips (1981) using the same data set found a significant negative relationship between teachers' verbal ability scores and students' Iowa Test of Basic Skills (ITBS) vocabulary score gains after controlling for teachers' experience and degree level, ratings of their undergraduate institutions, and students' socioeconomic status. These authors found even stronger positive relationships between teacher experience and student performance.

Similarly, when Summers and Wolfe (1977) considered the effects of teacher attributes on sixth graders' ITBS composite score gains, they found that although college selectivity ratings appeared significant, scores on the NTE Common Examinations—which measured general academic ability and included English, mathematics, social studies, and science components—were negatively related to students' achievement gains.

Stronger findings are provided by two studies at the school and district levels. (Walsh also cites these studies despite their presumed “aggregation bias.”) Ehrenberg and Brewer's (1995) reanalysis of data from the Equality of Educational Opportunity study (Coleman et al., 1966) examined the influence of several teacher characteristics on schools' average student gain scores while holding constant school, student, and community characteristics. They found that teachers' verbal aptitude scores were an important determinant of the school-to-school variation in student gain scores, with a smaller contribution of teachers' experience at the elementary level. Ferguson and Ladd (1996) found a positive, significant relationship between teachers' ACT college entrance examination scores and the achievement gains of third and fourth graders from Alabama in reading and mathematics in both school- and district-level analyses. Class size was an important

additional predictor at the school level, and teachers' degree level was an additional influence at the district level.

Although these studies suggest that teachers' verbal or general academic ability appears related to student achievement, none of them include measures of teacher education or certification. Thus, they cannot sustain a claim that verbal ability measures matter more than measures of teachers' professional knowledge. In fact, Ferguson and Ladd's (1996) findings in their Alabama study show smaller influences of teachers' ACT scores on student achievement (and greater influences of master's degrees) than Ferguson (1991) found for the Texas teacher licensing test, which includes components that come closer to evaluating knowledge that is used for teaching. In an article written nearly 20 years ago, Murnane (1983) observes that evidence about the influence of verbal ability was partly a function of the fact that such scores were among the few teacher variables available in large-scale data sets at that time. In his words,

Clearly one should not interpret these results as indicating that intellectual ability should be the sole criterion used in recruiting teachers or that formal teacher training cannot make a difference. In fact, the lack of evidence supporting formal preservice training as a source of competence may be to some extent a result of limitations in the available data. For example, all databases suitable for examining the correlates of teaching effectiveness as measured by student achievement gains pertain to a single school district. Since there is less variation in training among teachers within a district than among teachers in the country at large, these databases do not permit the most powerful possible tests of the efficacy of alternative teacher training programs. (p. 565)

Even strong advocates of the notion that academic ability matters are not willing to make the kinds of sweeping assertions found in the Secretary's report. For example, Eric Hanushek is quoted in the Secretary's report (p. 7) for his statement that "the closest thing to a consistent finding among the studies is that 'smarter' teachers who perform well on verbal ability tests do better in the classroom." The Secretary's report does not include Hanushek's next sentence, which reads, "Even for that, the

evidence is not very strong" (Hanushek, 1996, p. 116).

Subject Matter Knowledge

There is also evidence on the importance of subject matter knowledge to teaching. Rowan, Chiang, and Miller (1997) found that students who were taught by a teacher with a bachelor's or master's degree in mathematics or one who had scored well on a brief mathematics quiz had higher gains in achievement in this subject area, but that the effect was quite small—about 0.015 standard deviations in test score gains. In a study using data from NELS 1988, Goldhaber and Brewer (1997) report a greater influence on student achievement of teachers' bachelor's and master's degrees in the content area taught (e.g., mathematics or mathematics education) than was true for undifferentiated degrees.

A number of studies show the influences of subject matter knowledge in conjunction with knowledge about teaching. As noted earlier, Goldhaber and Brewer (2000) found a substantial influence of teachers' degrees in the content area or content area education on student achievement, alongside even larger effects of teacher certification. In a multilevel analysis of the LSAY data set, Monk and King (1994) report some evidence of cumulative effects of prior as well as proximate teachers' subject matter coursework on student performance in mathematics but did not find the same effects in science. Also as noted earlier, in another analysis of the same data set, Monk (1994) found that teachers' content preparation, as measured by coursework in the subject field, was usually positively though rarely significantly related to student achievement in mathematics and science, and that coursework in teaching methods had a stronger influence than additional coursework in mathematics. Monk's finding is reminiscent of Begle's (1979) finding from the National Longitudinal Study of Mathematical Abilities that teachers' coursework in mathematics methods had a stronger effect on student achievement than additional higher level coursework in mathematics for a group of already strong teachers. These studies do not suggest that subject matter knowledge is unimportant. However, they do call into question the Secretary's assertion that verbal ability and subject matter knowledge are more important

for teacher effectiveness than knowledge of how to teach.

Proposition 3: Teachers Who Have Completed Teacher Education Programs Are Academically Weak and Underprepared for Their Jobs

The Secretary's report also makes several misleading assertions regarding the qualifications of the teacher workforce. For example, in support of the assertion that "our system allows too many poorly qualified individuals into the classroom" (p. 12), the Secretary's report states that "only 38 percent [of teachers] have an undergraduate or graduate degree in an academic field outside of a school of education" (p. 12). However, the National Center for Education Statistics (NCES) data cited in the Secretary's report show that 95% of high school teachers and 66% of middle school teachers in 1998 had earned an academic degree in the subject area they were teaching or in subject area education (e.g., mathematics or mathematics education) (1999, p. 12). The Secretary may not have understood that candidates who complete a degree in science education, for example, have generally completed a content major or its equivalent plus additional education coursework. At many universities, a science education major requires as much or more science coursework than a regular major because candidates must fulfill distributional requirements across the sciences as well as in an area of concentration.

New teachers' levels of content preparation have improved since the 1980s, as 38 states now require a content major for teachers (U.S. Department of Education, 2002, p. 30). Fewer than half of all teachers now receive a bachelor's degree in education: Most complete another major and complete a minor, double major, or a credential in education or secure a master's degree. NCES data show that the proportion of high school teachers holding a major or minor in their main teaching field increased noticeably in all core academic fields between 1994 and 1998, reaching 90% or more in each area by 1998 (1999, pp. 19–20).

Requirements have also changed for elementary teachers. About 10 states require a subject area major or concentration. These states as well as those that expect degrees in elementary education or interdisciplinary

fields like liberal studies now require specific content courses across the curriculum that elementary teachers need to teach (National Association of State Directors of Teacher Education and Certification, 2001). These distributional requirements reverse the historical trend in which elementary teachers tended to have little or no coursework in fields like mathematics or science. Interestingly, Walsh's report (2001, p. 38) suggests that Maryland should eliminate content requirements for elementary teachers because many of the candidates for the state's short-term alternative route could not meet them.

The Secretary's report also asserts that "Research suggests that students enrolled in schools of education are not as academically accomplished as other university students" (p. 13). Quoting a study "conducted by *Education Week*" (p. 13) (in fact, it appears the study was conducted by NCES), the Secretary's report recounts the statement that "only 14 percent of the top quartile of 1992–93 college graduates entered some type of teacher-preparation program, only 12 percent actually taught, and a mere 11 percent stayed in the teaching profession through 1997." Because even smaller proportions of the top quartile of college graduates go into fields like medicine, law, and engineering, it is difficult to know what readers are intended to make of this statement. The retention rate implied by this analysis (11 of every 12 entrants stayed for 5 years) would be a good news story, suggesting retention rates for these prepared high-ability teachers of more than 90%. In fact, however, this appears to be a misstatement of the actual statistics. The real question about qualifications is reflected in the following claim:

Similar data from NCES also suggest that schools of education fail to attract the best students. For example, among college graduates who majored in education, just 14 percent had SAT or ACT scores in the top quartile, compared to 26 percent who majored in the social sciences and 37 percent who majored in mathematics, computer science, or the natural sciences. In contrast, 25 percent of uncertified teachers scored in the top quartile on these tests, as did 33 percent of private-school teachers. (2002, p. 13)

The NCES analysis, however, does not represent the range of training routes teachers now pursue because most college stu-

dents who prepare to teach no longer take an education major but complete a separate major and enroll in schools of education for a credential or master's degree. Looking at the full pool of candidates taking licensing examinations who prepared to teach or entered teaching (and using actual scores rather than self-reported data), the Educational Testing Service found that among 270,000 test takers in 1995 through 1997, the lowest pass rates on the Praxis II tests were experienced by individuals who had never enrolled in teacher education (Gitomer, Latham, & Ziomek, 1999). On SAT and ACT tests, those currently enrolled in teacher education slightly outscored Praxis takers who had never enrolled in teacher education. Although special education and physical education majors had SAT scores below the average for all college-bound seniors who took the SAT, elementary education majors did about as well, and teaching candidates in English, science, mathematics, social studies, and foreign language had higher mean SAT verbal and mathematics scores than all college-bound seniors who took the SAT. Scores for mathematics and science teachers were substantially higher than the overall pool of test takers on the math SAT, and scores for English, science, and foreign language teachers were substantially higher than the overall pool on the SAT verbal tests.

Finally, the Secretary's report erroneously asserts that "a majority of graduates of schools of education believe that traditional teacher preparation programs left them ill-prepared for the challenges and rigors of the classroom" (2002, p. 15), citing the following statistics:

According to NCES data, fewer than 36 percent of new teachers feel "very well prepared" to implement curriculum and performance standards, less than 30 percent feel prepared to integrate technology into instruction and less than 20 percent feel prepared to meet the needs of diverse students or those with limited English proficiency.

This is a very misleading statement. In fact, the NCES data cited in the Secretary's report are based on surveys that utilized a four-point scale; in the surveys, teachers were asked whether they felt "very well prepared," "moderately well prepared," "somewhat well prepared," or "not at all prepared" to carry out various activities in

the classroom. The survey results were not limited either to new teachers or to graduates of schools of education as the Secretary's report suggests. Instead, they are reported for all full-time public school teachers, including those who entered without preparation. Using the top two categories in the Likert scale, the data reveal that 95% of all teachers felt moderately or very well prepared to maintain order and discipline in the classroom; 82% felt adequately prepared to implement new methods of teaching; 77% of respondents felt prepared to implement curriculum and performance standards; 69% felt prepared to use student performance assessment techniques; 62% felt prepared to address the needs of students with disabilities; and 57% felt prepared to integrate technology into instruction. In each of these categories, fewer than 10% of teachers felt "not at all prepared." In the lowest category, only 53% felt very well or moderately well prepared to meet the needs of limited English proficient students and 17% of teacher felt not at all prepared—a finding that the NCES report notes reflects both the newness of this expectation and the fact that only 54% of teachers actually taught limited English proficient students (1999, p. 48). Of course, because this survey was reported for all teachers, it does not reveal the differences in feelings of preparation for teachers who experienced preparation of various kinds.

There are data on this latter point. In fact, several recent studies reveal that most teacher education graduates believe that their programs prepared them well for classroom teaching. For example, a survey of Kentucky teachers (Kentucky Institute for Educational Research, 1997) found that more than 80% of beginning teachers who graduated from Kentucky colleges of education felt well prepared for virtually all aspects of their jobs. Similarly, well over 70% of the graduates of the California State University felt well prepared for virtually all aspects of their jobs, and those who had student teaching (just over half of the total) felt significantly better prepared—and were viewed as better prepared by principals—than those who had completed certification through an internship program or who had taught on an emergency credential without student teaching (California State University, 2002a, 2002b). Finally, a 1998 survey of 3,000

beginning teachers in New York City found that teachers who were prepared in teacher education programs felt significantly better prepared for virtually all tasks of teaching than those who lacked preparation or entered teaching through alternative programs (Darling-Hammond, Chung, & Frelow, 2002).

Proposition 4: Alternative Certification Programs Have Academically Stronger Recruits, High Rates of Teacher Retention, and Produce More Successful Teachers

The section of the Secretary's report headed "Alternate Routes to Certification: A Model for the Future" is particularly replete with misinformation. The report claims that "performance on licensure tests is higher among alternate route teachers than traditionally trained teachers in most states" (p. viii) and later (p. 34) refers to a figure in the report for a similar assertion that alternate route pass rates are higher in 70% of states. However, the figure does not support this assertion: Summarizing state Title II self-reports on this issue, the figure is titled "Percent of states where those in alternative programs of teacher preparation have *equal or higher pass rates* on state assessments than those in traditional programs." The figure shows high rates of equivalent passage on basic skills and content tests (80–100%) and low rates on professional knowledge tests (45% of states). In addition to the fact that the figure does not support the Secretary's assertion, there are reasons to question what the data represented in the figure are measuring. The report later notes (p. 48) that a number of states did not report on their alternate routes and that California—a state earlier cited for its profusion of alternatives—indicated in its report that it did not have such routes. Furthermore, the report notes that pass rates in Massachusetts are 100%. However, a recent study on Massachusetts' alternative route program, Massachusetts Institute for New Teachers (MINT), found that 56% of MINT's 2002 recruits failed to pass the state content test and their scores were surpassed by test takers statewide on every test of content knowledge (Fowler, 2002).

As noted earlier, Gitomer, Latham, and Ziomek (1999) found that performance on the Educational Testing Service's Praxis tests is higher for those who have been en-

rolled in teacher preparation programs than those who did not participate in teacher education. Although this study does not specifically identify ACPs, it does indicate that trained teachers perform best on the most widely used licensure tests in the country. The authors note:

Current teacher education students have the highest passing rates on the licensure tests. Interestingly, those who report they have never been enrolled in a teacher education program have the lowest passing rates . . . These results make it clear that teacher education programs have an important impact in preparing their students to meet the requirements of licensure. (p. 24)

A possible source for the Secretary's assertion could be an appendix to Feistritzer and Chester's (2002) state-by-state catalogue of alternative routes across the country, which is cited elsewhere in the Secretary's report. This appendix quotes William Wale, Director of the Texas State Board for Educator Certification's Office of Educator Preparation, and includes some statistics showing pass rates on an unidentified teacher certification exam. These show slightly higher pass rates for ACP candidates in 1996–1997 than for traditional program candidates in Texas. However, another report about Texas's ACPs (Barnes, Salmon, & Wale, 1989) lists "alternative" programs ranging from university-based 5-year bachelor's plus master's degree models or preservice master's degree programs (which are called alternative because they are not undergraduate models) to district-run programs that place teachers in classrooms after a few weeks of summer training. Thus, these data make it impossible to draw any conclusions about the kinds of "alternatives" the Secretary favors (i.e., those that minimize education training).

Teacher Retention

In addition, the Secretary's report incorrectly asserts that

Initial evidence suggests that retention rates for teachers certified through alternate routes are higher than for teachers who enter the classroom through traditional routes. Nationwide, about 85 percent of teachers certified through alternate routes remain in the classroom five years later . . . (p. 16)

For this proposition, the report cites Feistritzer and Chester (2002). This com-

prehensive report is very helpful in its descriptions of state requirements for such routes, but it includes no citations to justify this claim. We surmise that the Secretary's sweeping claim may refer to an unreferenced statement on page 8 of that report, referring to California's intern teaching program and stating that "the retention for the first five years is 86 percent."

These data are neither national nor accurate for California. Two other versions of the retention statistic are offered in an appendix to Feistritzer and Chester's report, both referring to a paper by a California agency consultant (McKibbin, 1999) who is quoted as placing retention at 87% (p. 424) or 85% (p. 431) over 3 years, not five. This report's statistics are based in turn on an earlier report that studied a subset of recently funded California intern programs representing about one fourth of all such programs in the state, which cited a retention rate of about 85% for program graduates over the period of what appeared to be 1 year (McKibbin, 1998). The retention statistic is based on program self-reports rather than first-hand data and on program graduates rather than total program participants. An independent analysis of the data set examining all program participants indicated that about 80% of intern program participants (pregraduates) appeared to have remained in teaching after a year, although there was a substantial range across programs, and only about 60% remained by the 3rd year of teaching.

These data are comparable to other studies that have found relatively high attrition for interns during an initial year of teaching undertaken after a few weeks of preservice training. For example, an earlier evaluation of the Los Angeles Teacher Trainee program, California's largest district-run internship program, found that only 80.3% completed the 1st year of training and only 64.6% completed the 2nd year and received a clear credential the year after (Wright, McKibbin, & Walton, 1987). Another analysis of this same program revealed that 53% of the recruits had left the district within the first 5 years of program operation (Stoddart, 1992). Comparable attrition rates were found for an ACP in Dallas, Texas, which found only about 54% of recruits progressed from 1st year to 2nd year sta-

tus without “deficiencies” and only 40% planned to remain in teaching (Lutz & Hutton, 1989), and one in New York City (the Teaching Fellows Program), which lost more than 15% of its first class by Thanksgiving and more than 30% by the end of the initial year (Goodnough, 2000).

The Secretary’s report also states that in Texas, “retention rates are higher for African American teachers who entered teaching through an alternate route as compared to their traditionally certified teachers” (2002, p. 16). We noted that “alternative” routes in Texas include all postbaccalaureate programs, whether extended preservice or streamlined inservice programs. The Texas data included in Feistritz and Chester’s appendix show a 5-year retention rate between 1992 and 1997 of 72% for African-American ACP recruits versus a 66% rate for traditional recruits (2002, p. 419). However, the Secretary’s report neglects to note that drop out rates are also higher for African-American recruits from alternative programs (18% vs. 15% for traditional recruits). The Secretary’s report also neglects to note that these data show 5-year teacher retention rates for ACP recruits that are much lower than the 85% “national” retention figure cited elsewhere in the report and lower than those for traditional entrants (62% vs. 67%). Similarly, drop out rates for ACP recruits are higher than those for traditional entrants (29% vs. 25%) (Feistritz & Chester, p. 419).

Among ACPs, those that provide more extensive supervision and support both before and while recruits take over as independent teachers appear to have better outcomes in terms of retention. For example, New Haven, California has shown retention rates above 90% for its internship program with California State University—Hayward, which provides a full program of coherent coursework integrated with full-time student teaching followed by intensive mentoring when interns take over a part-time teaching load (Snyder, 1999). Similarly, a RAND report examining a national sample of programs found significantly higher rates of planned retention for recruits who graduated from alternative programs with more tightly coordinated and extensive preservice components and more intensive supervision than those who completed alternatives that featured only a few weeks of summer training before in-

dependent teaching (Darling-Hammond, Hudson, & Kirby, 1989).

The inclusion of student teaching appears to be one important element in predicting the outcomes of different programs. For example, a recent report from the NCES (2000) notes that 29% of newly graduated teachers who had not had student teaching left teaching within 5 years—an entry strategy that is typical of emergency hires and some of the shorter term alternative routes—as compared to only 15% of those who had had student teaching. However, the Secretary’s report suggests that student teaching, along with education school coursework, should be “optional” (p. 19) and urges that ACPs should not be “larded with a variety of requirements” (p. viii).

Effectiveness of Alternative Certification Programs

The Secretary’s report argues that, on the one hand, “traditional teacher training programs do not necessarily produce graduates with superior teaching skills while at the same time they impose significant costs and challenges on prospective teachers” and, on the other hand, “alternate routes to certification demonstrate that streamlined systems can boost the quantity of teachers while maintaining or even improving their quality” (p. 19). The report cites the findings of a study of Teach for America (TFA) recruits in Houston (Raymond, Fletcher, & Luque, 2001) as evidence for the success of alternative programs. The Secretary’s report states that “the evaluation reveals the district’s highest performing teachers are consistently TFA teachers, while the lowest performing teachers are consistently not TFA teachers” (2002, p. 18). This is not, however, what the study’s regression analyses actually found. The data reported by Raymond and colleagues show that experienced teachers in Houston were significantly more successful than inexperienced teachers, including TFA teachers. After controlling for teacher experience and school and classroom demographics, TFA recruits were found to be about as effective as other inexperienced teachers in schools and classrooms serving high percentages of minority and low-income students, which, the study shows, is where most underqualified teachers in the district were placed. In 1999–2000, the last year covered by the study sample,

about 50% of Houston’s new teachers were uncertified, and the researchers report that 35% of new hires lacked even a bachelor’s degree; so TFA teachers were compared to an extraordinarily ill-prepared group.

Because Houston’s distribution of underqualified teachers was intensely concentrated in schools serving the most disadvantaged students and because the regressions controlled for proportions of minority and low-income students at the school level as well as proportions of low-income and low-achieving students at the classroom level, TFA teachers were compared largely to the other underqualified teachers concentrated in these schools and classrooms. Although they have the data to do so, Raymond and colleagues do not report how TFA teachers’ outcomes compared to those of trained and certified teachers or to others with a bachelor’s degree.

The study indicates that minority students in Houston, who were disproportionately taught by these underprepared teachers, performed increasingly poorly each year in comparison with their White peers. The TFA study also reports, although the Secretary’s report does not, the extraordinarily high attrition rates for TFA teachers: Over the 3 years studied, from 60 to 100% of TFA recruits had left after their 2nd year of teaching.

A recent study in five Arizona school districts did examine the relative effectiveness of TFA teachers as compared to other new teachers with different levels of qualifications. Using a matched comparison design in which teachers were matched by experience level, grade level taught, level of education, and school or district, the study found that the students of uncertified teachers, including TFA teachers, did significantly poorer than those of comparably experienced certified teachers on mathematics, reading, and language arts tests (Laczko-Kerr & Berliner, 2002). Another recent study found that TFA recruits in New York City, like other untrained recruits there, felt markedly less well prepared for teaching than graduates of teacher education programs. These less well-prepared entrants to teaching reported that they felt significantly less efficacious in their teaching and less able to meet their students’ needs, less satisfied with their training, less likely to stay in teaching, and less likely to say they would come into teaching through the same pathway again (Darling-Hammond,

Chung, & Frelow, 2002). Although each of the three TFA-related studies has limitations, in combination they illustrate that no sweeping claims can be made for the effectiveness of the program.

There is other research on alternative routes that goes beyond the TFA program, which, as Feistritzer and Chester (2002) accurately note, is not actually an ACP because its recruits teach on emergency permits and are not part of a coherent program of teacher education (which most states with alternative certification require). This research is very difficult to interpret because the design and quality of programs labeled traditional and alternative varies greatly both within and across states, with some states labeling as alternative postbaccalaureate programs that other states would call traditional. Some alternate route programs provide scant preparation while others involve extensive coursework, pre- and in-service professional development, and school- and university-based induction support. Furthermore, many studies do not report on the design features of the programs they have examined.

Thus, it may not be surprising that the findings on ACPs are mixed: A number of studies have found that ratings of the competence of alternative program candidates by principals and supervisors are more frequently negative than those of “traditional” recruits (see Gomez & Grobe, 1990 regarding Dallas; Jelmberg, 1996 regarding New Hampshire; Mitchell, 1987 regarding Dallas; CSU, 2002a, 2002b regarding internship programs in California), while others have found that ratings are comparable for candidates from alternative and traditional programs (Miller, McKenna, & McKenna, 1998). Only two controlled studies of the achievement outcomes of ACP and traditionally trained teachers have been reported, with one finding the students of traditional teachers showed significantly larger gains in language arts than those of ACP teachers (Gomez & Grobe, 1990), and the other finding student outcomes comparable across the two groups, although gain scores were not reported (Miller, McKenna, & McKenna, 1998).

When this research is analyzed in terms of program design, it appears that more carefully designed programs yield stronger outcomes in terms of teacher effectiveness and retention than those that provide less training and support. For example, Jelm-

berg (1996) compared traditionally certified teachers in New Hampshire with teachers in the same state who had participated in a state-sponsored ACP. The alternate-route teachers assumed “full responsibility for students prior to any preparation, and (had) three years” to acquire 14 state-identified competencies through workshops or college courses (p. 61). Based on surveys of 136 principals and more than 200 teachers, the author reports that traditionally certified teachers were rated by their principals significantly higher than alternate-route teachers on instructional skills and instructional planning, and the traditionally certified teachers rated their own preparation significantly higher than did the alternatively certified teachers.

By contrast, Miller, McKenna, and McKenna (1998) found no differences between traditionally certified and alternately certified teachers with regard to instructional practices or student achievement in a university-sponsored program. This program offered 15 to 25 credit hours of coursework before interns entered classrooms where they were intensively supervised and assisted by both university supervisors and school-based mentors while they completed additional coursework needed to meet full standard state certification requirements. Because the design of this program was so different from many quick-entry alternative route programs, Miller, McKenna, and McKenna (1998) note that their studies

provide no solace for those who believe that anyone with a bachelor’s degree can be placed in a classroom and expect to be equally successful as those having completed traditional education programs . . . The three studies reported here support carefully constructed AC programs with extensive mentoring components, post-graduation training, regular in-service classes, and ongoing university supervision. (p. 174)

Ironically, this more extensive education coursework and more heavily mentored clinical training compose the very “bureaucratic hurdles” the Secretary’s report suggests should be eliminated (p. 19).

Conclusion

Having asserted that requirements regarding education coursework and student teaching are “the Achilles heel of the certification system” (p. 31), the Secretary’s re-

port concludes its peculiar rendering of the research on teaching with the following recommendation: “To meet the ‘highly qualified’ teachers challenge, then, states will need to streamline their certification system to focus on the few things that really matter: verbal ability, content knowledge, and, as a safety precaution, a background check of new teachers” (p. 40).

As our review indicates, these assertions and policy recommendations are not supported by scientifically based research. Although there is evidence that verbal ability and content knowledge contribute to teacher effectiveness, there is also evidence that teacher preparation—including the student teaching and methods coursework the Secretary’s report deplores—contributes at least as much to outcomes ranging from teacher effectiveness to teacher retention. And although there is evidence that some well-designed ACPs have strong outcomes, there is also evidence that programs and entry pathways that skirt the core features of teacher preparation produce recruits who consider themselves underprepared, are viewed as less competent by principals, are less effective with students, and have high rates of teacher attrition. Finally, the recent advances states have made in strengthening teacher certification requirements have begun to be evident in stronger academic backgrounds and licensing test scores for college graduates who have prepared to teach. These trends suggest that meeting the highly qualified teacher challenge will require states to stay the course with respect to the gains they have already made, rather than to reverse course on the basis of a fictionalized account of what research says about what effective teachers know and how they come to know it.

NOTES

¹ Among these are the Fordham Foundation, which has issued a “manifesto” urging the elimination of teacher education and certification requirements and the Abell Foundation, which has advocated for similar measures.

² *Teacher Certification Reconsidered: Stumbling for Quality*, sponsored by the Abell Foundation, is published on the Abell Foundation website at www.abellfoundation.org. The version of the report that was publicized and published on this website in September 2001 is the basis for this response. Walsh states that some of the earlier errors and misrepresentations noted in a reply to that report (Darling-Hammond,

2001) have been removed from a version since published in hard copy by the Foundation in December 2001 (Walsh & Podgursky, 2001).

³ The versions of the report published on the Foundation's website have changed as it has been critiqued. These citations were in the versions made available in August and September 2001 and were removed in the version made available on the Abell Foundation website in March 2002.

⁴ Most states require a specified amount of student teaching, and a growing number require a probationary period, which may feature additional mentoring and assessment before full certification can be awarded.

⁵ The Secretary's report references an article about the study published in the Fordham Foundation report (Goldhaber & Brewer, 1999) before the study was published in a peer-reviewed journal (Goldhaber & Brewer, 2000).

⁶ In states with the most stringent certification laws, a teacher fully certified in chemistry but not in physics may be teaching on a regular license in one subject and a temporary or emergency license in the other, which is why temporary certification is most common in science.

REFERENCES

- Barnes, S., Salmon, J., & Wale, W. (1989). *Alternative teacher certification in Texas*. Paper presented at the annual meeting of the American Educational Research Association. (ERIC Document No. 307316)
- Begle, E. G. (1979). *Critical variables in mathematics education: Findings from a survey of empirical literature*. Washington, DC: National Council for Teachers of Mathematics.
- Betts, J. R., Rueben, K. S., & Danenberg, A. (2000). *Equal resources, equal outcomes? The distribution of school resources and student achievement in California*. San Francisco: Public Policy Institute of California.
- Bowles, S., & Levin, H. M. (1968). The determinants of scholastic achievement—An appraisal of some recent evidence. *Journal of Human Resources*, 3, 3–24.
- Bullough, R., Knowles, J., & Crow, N. (1989). Teacher self-concept and student culture in the first year. *Teachers College Record*, 91(2), 209–233.
- California State University. (2002a). *First system wide evaluation of teacher education programs in the California State University: Summary report*. Long Beach, CA: Office of the Chancellor, California State University.
- California State University. (2002b). *Preparing teachers for reading instruction (K–12): An evaluation brief by the California State University*. Long Beach, CA: Office of the Chancellor, California State University.
- Chandler, W., Jelacic, S., Boster, D., Ciol, M., Williams, G., Watkins, S., et al. (2002). Prothrombotic coagulation abnormalities preceding the hemolytic-uremic syndrome. *The New England Journal of Medicine*, 346(1), 23–32.
- Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., et al. (1966). *Equality of educational opportunity*. Washington, DC: U.S. Government Printing Office.
- Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Educational Policy Analysis Archives*, 8(1). Retrieved from <http://epaa.asu.edu/epaa/v8n1>
- Darling-Hammond, L. (2001). The research and rhetoric on teacher certification: A response to "Teacher certification reconsidered." New York: National Commission on Teaching and America's Future. Retrieved from http://www.nctaf.org/publications/abell_response.pdf
- Darling-Hammond, L. (2002). The research and rhetoric on teacher certification: A response to "Teacher certification reconsidered." *Educational Policy Analysis Archives*, 10(36). Retrieved from <http://epaa.asu.edu/epaa/v10n36>
- Darling-Hammond, L., Berry, B., & Thoreson, A. (2001). Does teacher certification matter? Evaluating the evidence. *Educational Evaluation and Policy Analysis*, 23(1), 57–77.
- Darling-Hammond, L., Chung, R., & Frelow, F. (2002). Variation in teacher preparation: How well do different pathways prepare teachers to teach? *Journal of Teacher Education*, 53(4), 286–302.
- Darling-Hammond, L., Hudson, L., & Kirby, S. (1989). *Redesigning teacher education: Opening the door for new recruits to science and mathematics teaching*. Santa Monica, CA: The RAND Corporation.
- Ehrenberg, R. G., & Brewer, D. J. (1995). Did teachers' verbal ability and race matter in the 1960s? Coleman revisited. *Economics of Education Review*, 14(1), 1–21.
- Feistritzer, C. E., & Chester, D. T. (2002). *Alternative teacher certification: A state-by-state analysis 2002*. Washington, DC: National Center for Education Information.
- Ferguson, R. F. (1991). Paying for public education: New evidence on how and why money matters. *Harvard Journal on Legislation*, 28(2), 465–498.
- Ferguson, R. F., & Ladd, H. F. (1996). How and why money matters: An analysis of Alabama schools. In H. F. Ladd (Ed.), *Holding schools accountable: Performance-based reform in education* (pp.265–298). Washington, DC: Brookings Institution.
- Fetler, M. (1999). High school staff characteristics and mathematics test results. *Educational Policy Analysis Archives*, 7(9). Retrieved from <http://epaa.asu.edu/epaa/v7n9>
- Fowler, C. (2002). Fast track . . . Slow going? Research Brief. *Ed Policy*, 2(1). Retrieved September 2002 from <http://www.edpolicy.org/publications/documents/updatev2i1.pdf>
- Fuller, E. (1998). *Do properly certified teachers matter? A comparison of elementary school performance on the TAAS in 1997 between schools with high and low percentages of properly certified regular education teachers*. Austin: The Charles A. Dana Center, University of Texas at Austin.
- Fuller, E. (2000). *Do properly certified teachers matter? Properly certified Algebra teachers and Algebra I achievement in Texas*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- Gitomer, D. H., Latham, A. S., & Ziomek, R. (1999). *The academic quality of prospective teachers: The impact of admissions and licensure testing*. Princeton, NJ: Educational Testing Service.
- Goe, L. (2002, October 14). Legislating equity: The distribution of emergency permit teachers in California. *Educational Policy Analysis Archives*, 10(42). Retrieved November 8, 2002 from <http://epaa.asu.edu/epaa/v10n42>.
- Goldhaber, D. D., & Brewer, D. J. (1997). Evaluating the effect of teacher degree level on educational performance. In W. J. Fowler (Ed.), *Developments in School Finance, 1996* (pp.197–210). Washington, DC: National Center for Education Statistics, U.S. Department of Education.
- Goldhaber, D. D., & Brewer, D. J. (1999). Teacher licensing and student achievement. In C. E. Finn & M. Kanstoroom (Eds.), *Better teachers, better schools*. Washington, DC: Thomas B. Fordham Foundation. Retrieved from <http://www.edexcellence.net/better/goldhab.pdf>
- Goldhaber, D. D., & Brewer, D. J. (2000). Does teacher certification matter? High school teacher certification status and student achievement. *Educational Evaluation and Policy Analysis*, 22(2), 129–146.
- Gomez, D. L., & Grobe, R. P. (1990). *Three years of alternative certification in Dallas: Where are we?* Paper presented at the annual meeting of the American Educational Research Association, Boston, MA.
- Goodnough, A. (2000, November 22). Winning process begins for novice teachers, *New York Times*. Retrieved from <http://nytimes.com>.
- Hanushek, E. A. (1992). The trade-off between child quantity and quality. *Journal of Political Economy*, 100(1), 84–117.
- Hanushek, E. A. (1996). *School resources and achievement in Maryland*. Baltimore: Maryland State Department of Education.
- Hanushek, E. A., Rivkin, S. G., & Taylor, L. L. (1995). *Aggregation bias and the estimated effects of school resources*. Rochester, NY: University of Rochester, Center for Economic Research.

- Hawk, P. P., Coble, C. R., & Swanson, M. (1985). Certification: It does matter. *Journal of Teacher Education, 36*(3), 13–15.
- Jelmborg, J. (1996). College-based teacher education versus state-sponsored alternative programs. *Journal of Teacher Education, 47*(1), 60–66.
- Kentucky Institute for Education Research. (1997). *The preparation of teachers for Kentucky schools: A survey of new teachers*. Frankfort: Kentucky Institute for Education Research.
- Laczko-Kerr, I., & Berliner, D. (2002). The effectiveness of Teach for America and other under-certified teachers on student academic achievement: A case of harmful public policy. *Educational Policy Analysis Archives, 10*(37). Retrieved from <http://epaa.asu.edu/epaa/v10n37>
- Lutz, F. W., & Hutton, J. B. (1989). Alternative teacher certification: Its policy implications for classroom and personnel practice. *Educational Evaluation and Policy Analysis, 11*(3), 237–254.
- Massey, H. & Vineyard, E. (1958). Relationship between scholarship and first year teaching success. *Journal of Teacher Education, 9*, 297–301.
- McKibbin, M. (1998). *Teaching internship programs: Alternative preparation and licensure in California: Purposes, procedures and performance*. Sacramento: California Commission on Teacher Credentialing.
- McKibbin, M. (1999). *A report on teaching internship grant programs, 1994–1999: Lessons learned and challenges to face*. Sacramento: California Commission on Teacher Credentialing.
- Miller, J. W., McKenna, M. C., & McKenna, B. A. (1998). A comparison of alternatively and traditionally prepared teachers. *Journal of Teacher Education, 49*(3), 165–176.
- Mitchell, N. (1987). *Interim evaluation report of the alternative certification program (REA87-027-2)*. Dallas, TX: DISD Department of Planning, Evaluation, and Testing.
- Monk, D. H. (1994). Subject area preparation of secondary mathematics and science teachers and student achievement. *Economics of Education Review, 13*(2), 125–145.
- Monk, D. H., & King, J. A. (1994). Multilevel teacher resource effects in pupil performance in secondary mathematics and science: The case of teacher subject matter preparation. In R. G. Ehrenberg (Ed.), *Choices and consequences: Contemporary policy issues in education* (pp. 29–58). Ithaca, NY: ILR Press.
- Murnane, R. J. (1983). Understanding the sources of teaching competence: Choices, skills, and the limits of training. *Teachers College Record, 84*(3), 564–589.
- Murnane, R. J., & Phillips, B. R. (1981). What do effective teachers of inner-city children have in common? *Social Science Research, 10*, 83–100.
- National Association of State Directors of Teacher Education and Certification. (2001). *The NASDTEC manual on the preparation and certification of educational personnel*. Dubuque, IA: Kendall Hunt Publishing.
- National Center for Education Statistics. (1999). *Teacher quality: A report on the preparation and qualifications of public school teachers (NCES 1999-080)*. Washington, DC: U.S. Department of Education.
- National Center for Education Statistics. (2000). *Digest of education statistics, 1999*. Washington, DC: U.S. Department of Education.
- Quaini, F., Urbanek, K., Beltrami, A., Finato, N. L., Beltrami, C., Nadal-Ginard, B., et al. (2002). Chimerism of the transplanted heart. *The New England Journal of Medicine, 346*(1), 5–15.
- Raymond, M., Fletcher, S. H., & Luque, J. (2001). *Teach for America: An evaluation of teacher differences and student outcomes in Houston, Texas*. Stanford, CA: The Hoover Institution, Center for Research on Education Outcomes.
- Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2001). *Teachers, schools, and academic achievement*. Amherst, MA: Amherst College.
- Rowan, B., Chiang, F., & Miller, R. J. (1997). Using research on employees' performance to study the effects of teachers on students' achievement. *Sociology of Education, 70*(4), 256–284.
- Sanders, W. L., & Horn, S. P. (1994). The Tennessee Value-Added Assessment System (TVAAS): Mixed-model methodology in educational assessment. *Journal of Personnel Evaluation in Education, 8*(3), 299–311.
- Sanders, W. L., & Rivers, J. C. (1996). *Cumulative and residual effects of teachers on future student academic achievement*. Knoxville: University of Tennessee Value-Added Research and Assessment Center.
- Skjoerven, R., Wilcox, A., & Lie, R. (2002). The interval between pregnancies and the risk of preeclampsia. *The New England Journal of Medicine, 346*(1), 33–38.
- Snyder, J. (1999). *New Haven Unified School District: A teaching quality system for excellence and equity*. New York: National Commission on Teaching and America's Future.
- Stoddart, T. (1992). An alternate route to teacher certification: Preliminary findings from the Los Angeles Unified School District Intern Program. *Peabody Journal of Education, 67*, 3.
- Strauss, R. P., & Sawyer, E. A. (1986). Some new evidence on teacher and student competencies. *Economics of Education Review, 5*(1), 41–48.
- Summers, A. A., & Wolfe, B. L. (1975). *Equality of educational opportunity quantified: A production function approach* (Research paper 6). Philadelphia: Federal Reserve Bank of Philadelphia, Department of Research.
- Summers, A. A., & Wolfe, B. L. (1977). Do schools make a difference? *American Economic Review, 67*, 639–652.
- U.S. Department of Education. (2002). *Meeting the highly qualified teachers challenge: The Secretary's annual report on teacher quality*. Washington, DC: U.S. Department of Education, Office of Postsecondary Education, Office of Policy, Planning, and Innovation.
- Walsh, K. (2001). *Teacher certification reconsidered: Stumbling for quality*. Baltimore, MD: Abell Foundation. Retrieved September 2001 from http://www.abell.org/pubsitems/ed_cert_1101.pdf
- Walsh, K., & Podgursky, M. (2001). *Teacher certification reconsidered: Stumbling for quality. A rejoinder*. Baltimore, MD: Abell Foundation. Retrieved from http://www.abell.org/pubsitems/ed_cert_rejoinder_1101.pdf
- Wayne, A. J., & Youngs, P. (in press). Teacher characteristics and student achievement gains: A review. *Review of Educational Research*.
- Wenglinsky, H. (2000). *How teaching matters: Bringing the classroom back into discussions of teacher quality*. Princeton, NJ: Educational Testing Service.
- Wilson, S. M., Floden, R., & Ferrini-Mundy, J. (2001). *Teacher preparation research: Current knowledge, gaps, and recommendations. A research report prepared for the U.S. Department of Education*. Seattle: Center for the Study of Teaching and Policy, University of Washington.
- Wright, D. P., McKibbin, M. D., & Walton, P. H. (1987). *The effectiveness of the Teacher Trainee Program: An alternate route into teaching in California*. Sacramento: California Commission on Teacher Credentialing.
- Wright, S. P., Horn, S. P., & Sanders, W. L. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education, 11*, 57–67.

AUTHORS

LINDA DARLING-HAMMOND is the Charles E. Ducommun professor of education at Stanford University. Her research interests focus on educational policy and practice in the areas of teacher quality, school restructuring, and educational equity.

PETER YOUNGS is a postdoctoral fellow and research associate at Stanford University. He studies policy related to teacher education, licensure, and induction.

Manuscript received August 29, 2002

Revisions received October 8, 2002

Accepted October 9, 2002