

LINDA DARLING-HAMMOND

**Reforming Teacher Preparation and Licensing:
Debating the Evidence**

Teachers College Record 102 no1 28-56 F 2000

ABSTRACT

This article responds to Dale Ballou and Michael Podgursky's claims that the National Commission on Teaching and America's Future has misrepresented research data and findings. After reviewing and responding to each of their charges, the article indicates the ways in which their critique itself has misreported data and misrepresented the Commission's statements and recommendations. Ballou and Podgursky ignore and misconstrue the research evidence presented by the Commission in support of its key conclusions. Following an analysis of the ways in which the critique misrepresents the findings from research on teacher education to bolster the argument that training for teaching is unnecessary, this reply offers an argument for professional teaching standards as a key factor in achieving greater equity and excellence in American schools.

What Matters Most, the 1996 report of the National Commission on Teaching and America's Future (the Commission), was the

product of a 26-member bipartisan panel of governors, legislators, business leaders, community leaders, and educators. The Commission was concerned with understanding what it would take to enable every child in America to reach the new high standards of learning being enacted by states across the nation. Following two years of intensive study and debate, the Commission concluded that recent reforms like new curriculum standards, tests, and accountability schemes are unlikely to succeed without major investments in teaching. The history of school reform has illustrated that innovations pursued without adequate investments in teacher training have failed time and again. Furthermore, the demands of new subject matter standards and much more diverse student bodies require deeper content knowledge, more sophisticated pedagogical and diagnostic skills, and broader repertoires of teaching strategies for teachers.

The Commission report offered an interlocking array of 22 recommendations aimed at ensuring "a caring, competent, and qualified teacher" for every child, working in schools organized to support their success. The Commission's recommendations have stimulated legislative initiatives to improve teaching in more than 25 states, in many more local districts, and at the federal level. The recommendations also provide the basis for ongoing work with 15 state partners, more than 40 national partner organizations, and 9 urban districts that are working together to develop more effective systems for recruiting, preparing, inducting, evaluating, and supporting teachers and for redesigning schools so that they can support more powerful learning for a wider range of students.

As the Commission's implementation efforts have unfolded, attacks on its work have been circulated by Dale Ballou and Michael Podgursky through the Government Union Review (1997a) and Chester Finn's Thomas B. Fordham Foundation (1999). The major substantive points of disagreement revolve around two issues: Ballou and Podgursky, along with Finn and other contributors to the Fordham Foundation (1999) "manifesto," object to standards for teaching or teacher education that would influence the knowledge and skills teachers are expected to acquire or that schools of education might be expected to impart. Instead they prefer that administrators select teachers from the open market and evaluate them according to students' test scores.

In addition, they object to the idea that school resources might contribute as much to student achievement as student background characteristics such as race or parent income, and that states therefore have an obligation to equalize access to those resources. They are content to allow the market to operate unrestrained in the face of inequities in access to dollars, qualified teachers, curriculum, and class sizes that contribute to disparate outcomes between advantaged and disadvantaged students. Their proposed solutions would continue and exacerbate these disparities by creating additional disincentives for the most able teachers to work with the most needy students, by removing requirements for even minimally equitable treatment of less advantaged students in the allocation of teachers and other resources, and by reducing teachers' access to knowledge that might help them to become more effective.

In past critiques, Ballou and Podgursky have argued that the

Commission report was the work of education insiders, ungrounded in research, and likely to shift control from governments to private organizations. In this volume of the Teachers College Record, Ballou and Podgursky (1999a) go further to charge, falsely in each instance, that the Commission has misrepresented research data and findings. In the course of their argument, their critique itself misreports data, misrepresents the Commission's statements and recommendations, and variously ignores and misconstrues the research evidence presented in support of the report's key findings. In addition, the authors misrepresent the findings of research on teacher education in an effort to argue that training for teaching is unnecessary. This reply offers evidence on the key points of contention.

BALLOU AND PODGURSKY'S MISREPRESENTATIONS

The commission's membership and recommendations. In the first pages of their article, Ballou and Podgursky mischaracterize the membership of the Commission, describing it as "representatives of various education interest groups, including the presidents of the National Education Association, the American Federation of Teachers, the National Council for the Accreditation of Teacher Education (NCATE), and the National Board for Professional Teaching Standards (NBPTS)." This statement completely ignores the other 22 members of the Commission, including its Democratic and Republican governors, two state legislators, a current and a former state superintendent, three business leaders, two community leaders, two college presidents, two leaders of school reform initiatives, an education dean and a professor, a local

superintendent, principal, and three classroom teachers. Despite the breadth of the group as a whole, because these four education organizations were represented, Ballou and Podgursky (1997) have suggested elsewhere that the Commission was uncritically supportive of unions and of professional certification and accreditation.

Ballou and Podgursky also ignore 19 of the Commission's 22 recommendations, including those that address student standards; new models of teacher preparation, induction, and professional development; recruitment reforms, including streamlined hiring procedures, reciprocity in licensing, and alternative pathways for mid-career entrants; rewards for teacher knowledge and skill, including recognition of expert teachers and strategies to remove incompetent teachers; and redesigned schools that allocate more resources to the front lines of teaching.

The recommendations emerged from a 2-year process of research and vociferous debate in which each had to pass muster with the entire group of diverse, strong-willed commissioners, which was no easy task. Ballou and Podgursky's (1997a) allegations that the Commission offered a narrow agenda controlled by the education establishment ignores the fact that the report included recommendations that have previously been opposed by the same groups they suggest controlled the agenda, including changes in union policies that protect incompetent teachers and university practices that undermine teacher quality. Ballou and Podgursky's one-sided treatment of the Commission's proposals reflects the ideological lens they apply to the work.

The definition of teacher expertise. Ballou and Podgursky accuse the Commission of misrepresenting research findings because we use the term teacher "expertise" to discuss a group of variables together (e.g., teacher test scores, master's degrees, and experience). They seem to be suggesting that the only variable that should properly be reported as a measure of expertise is the percentage of teachers with master's degrees, and they discuss the issue as though the Commission's major recommendation regarding the preparation of teachers is to require master's degrees. They then look at large-scale quantitative studies for support for the idea that master's degrees are a major determinant of student achievement.

Their argument makes little sense. First of all, the notion that master's degrees are the best or only measure of expertise is one that exists only in Ballou and Podgursky's imagination. It is not a notion put forth by the Commission, which recommends a comprehensive array of measures to improve teachers' qualifications and access to knowledge, including policies and practices influencing admissions to teacher education; college studies in the liberal arts, subject matter, and education; extended student teaching or internships in professional development schools; licensing tests of both subject matter knowledge and teaching skills; induction for beginning teachers; and ongoing professional development.

While the Commission spoke approvingly of new 5-year teacher education models that have produced better entry and retention rates and evidence of effectiveness (see below) and recommended that teacher education extend beyond the typical 4-year

undergraduate program to provide time for a year-long internship, it did not suggest that states require master's degrees for teaching, either at the preservice or inservice levels. In fact, several of the programs the Commission described as high quality (including 5th-year models) do not require a master's degree. The goal is to construct an interlocking set of policies that encourage high-ability individuals into teaching, provide them access to knowledge that will improve their effectiveness, and keep them in the classroom so that their knowledge and experience become a resource for children's learning.

Second, research on teaching has developed a view of teaching expertise that includes general knowledge and ability, verbal ability, and subject matter knowledge as foundations; abilities to plan, organize, and implement complex tasks as additional factors; knowledge of teaching, learning, and children as critical for translating ideas into useful learning experiences; and experience as a basis for aggregating and applying knowledge in nonroutine situations. David Berliner's studies of expertise in teaching, for example, include experience along with several of these other traits as a critical aspect of expertise (see, e.g., Berliner, 1986). All of these factors combine to make teachers effective; furthermore, one cannot fully partial out the effects of one factor as opposed to another as many are highly correlated. It is because expertise is a product of many kinds of knowledge and skill that the Commission recommended a balanced, multifaceted preparation for teaching.

Third, given the wide variability in the content of master's degrees pursued by teachers, it would be difficult to describe them as

representing any common body of knowledge or "treatment." Few teachers have pursued master's degrees through a coherent program of preservice education tightly linking content, content pedagogy, and clinical training, like the 5-year programs the Commission described. Some master's degrees have been directly related to teaching and represent a coherent program of studies, for example, those in reading or special education most commonly taken by elementary teachers. (Interestingly, many studies find a larger effect of master's degrees on student achievement at the elementary level and in reading, although one should not make too much of this fact, given the points discussed below.) Other master's degrees have been a collection of courses that may or may not address teachers' abilities in the classroom to a great extent. Many have been pointed at jobs outside of teaching, such as administration, guidance counseling, measurement and evaluation, and the like. Thus, there is reason to expect that some master's degree studies would affect teaching ability, but not much reason to expect the effect of master's degrees as an undifferentiated variable to be uniform or large in the aggregate. The Commission suggested that master's degrees for which salary credit is awarded be more directly focused on improving teachers' classroom teaching knowledge and skills, in part because they have not always been in the past.

Finally, it would be particularly odd to find a large effect of master's degrees in a multivariate study of the sort Ballou and Podgursky focus on here. For one thing, variables describing teacher expertise are highly correlated; teachers with higher test

scores and those with greater experience are much more likely to have a master's degree. Other variables frequently examined in large scale studies, such as certification status, are also correlated with master's degrees. As Ballou and Podgursky understand, variable collinearity means that when two or more highly correlated variables are entered into an equation, one absorbs some of the variance that might otherwise appear to be explained by another. Thus, one cannot disaggregate with precision the effects of one such variable from those of another that is closely related to it. Their suggestion that the Commission should have done this in order to "prove" the effects of master's degrees--a single crude proxy for a wide array of knowledge and skills--contradicts their admonitions elsewhere in the paper.

Ferguson's Texas study. Ballou and Podgursky object to the Commission's description of Ronald Ferguson's (1991) study of 900 Texas school districts, which found a large influence of teacher qualifications on student achievement at the district level. First, as described above, they object to the Commission's labeling as "teacher expertise" the set of variables that include teachers' scores on the Texas licensing test (TECAT), teacher education (master's degrees), and teacher experience. We believe that all of these variables and others may capture aspects of teacher expertise that interact with one another, and we do not hold special warrant for master's degrees as the central indicator of expertise. In studies using large quantitative data sets in which all of the variables are crude proxies for more nuanced conditions, including many that are unmeasured, one can only speculate about the finer grained

phenomena that are at work in the real world. Our point in presenting the Ferguson findings was not to argue for master's degrees in education, or for any specific teacher education approach (we address these ideas elsewhere in the Commission's reports using different evidence bases, described later in this article), but to argue that policies that aim to enhance teachers' abilities and the equitable distribution of well-qualified teachers are important.

Ballou and Podgursky note correctly that the TECAT scores, given the other variables in the equations, represent the lion's share of the variance attributable to teachers' characteristics in Ferguson's estimates. Having described the TECAT as a test of basic literacy, they want to argue that the only implication of Ferguson's findings is that it may be important to have literate teachers, and "is there anyone who thinks otherwise?" Clearly policy makers who constructed a system of teacher hiring in Texas that did not distribute teachers equitably on this or other measures must have thought otherwise. That is the point of the Commission's recommendations and the import of Ferguson's findings: that teacher qualifications matter in policy decisions about resource allocations and school improvement.

A critical policy question is what Ferguson's finding, necessarily based on an easily available, large-scale quantitative measure, actually represents. The TECAT is actually a somewhat broader test than Ballou and Podgursky indicate, as it tests verbal ability, logical thinking and research skills, and professional knowledge. Knowing the content of the test helps only a bit in drawing inferences from these findings, however. A study of this kind can indicate only that

this measure of teachers' general and professional knowledge and skills--and whatever unmeasured attributes it may correlate with--is strongly associated with average student achievement. As Ferguson notes:

We can only speculate about what teachers with high scores do differently from teachers with low scores. One possibility is that teachers with higher scores more often attended colleges that were effective at preparing them to become good teachers. Alternatively, or in addition, teachers with higher scores may have thinking habits that make them more careful in preparation of lesson plans or more articulate in oral presentation. (p. 477)

Other kinds of research are needed (some are cited in the Commission's report; others are suggested elsewhere; see, e.g., Darling-Hammond, 1998) to figure out what programs and policies are likely to make a meaningful difference in the kinds of knowledge and skills that have been found in finer grained studies to be important for effective teaching.

In addition to their own narrow reading of the term teacher expertise, Ballou and Podgursky want to take issue with Ferguson's finding that school inputs, including teachers, may account for as much or more of the variance in student achievement among districts as do student background variables. They acknowledge that such a finding is "extraordinary," and note, "It appears to controvert a long-standing belief among researchers, dating back to the work of James Coleman in the 1960s, that students' family backgrounds

have far more explanatory power as predictors of achievement than schooling variables."

Ballou and Podgursky look for ways to buttress this long-standing belief. Calling the Commission's account of the research "incorrect and misleading," they first complain about the level of aggregation of the data. Their observation that "the same teacher qualifications may account for a very different share of explained variation in individual student achievement" than they do when assessing interdistrict variation is obvious and uncontested. The Commission accurately noted in its summary of the findings that this was a district level study. The research studied what it studied with the data that were available and produced interesting and important findings. The fact that Ballou and Podgursky do not like the findings because they challenge presumptions about school inputs and student backgrounds does not make them invalid or their presentation misleading. Ironically, later in their paper Ballou and Podgursky try to establish even more sweeping claims about the relationships between policies and student achievement using student achievement data aggregated to the state level with a bizarre study that includes independent variables that do not measure teacher qualifications and that postdate their achievement data. (We discuss this strange analysis later.)

Ballou and Podgursky take issue with the Commission's statement that in the study's various estimates of influences on average student achievement, teachers' expertise "accounted for about 40% of the measured variance ... more than any other single factor." The 40% statistic was derived by summing the percentages offered in the

article for variance explained by teachers' scores, teacher experience, and master's degrees, and by examining published and unpublished equations showing the combined variables accounting for between 38% and 43% of the variance, after taking account of the set of other controls in the estimates. While Ballou and Podgursky are technically correct that issues of covariance make it impossible to precisely assign proportions of variance to variables as though they are wholly orthogonal, our procedure was not unreasonable and the estimate not far afield as a simple summary to enable readers to gauge the relative influence of these variables compared to others measured. The point we correctly made is that this study demonstrates that teachers are very important--at least comparable in importance to parents or other family background factors when making comparisons across districts.

While this finding may be "extraordinary" in light of the presumption that school inputs don't matter, it is not unique. Another study using licensing test scores (Strauss & Sawyer, 1986) found similarly strong influences of teacher quality (as measured by teachers' scores on the National Teacher Examinations that measure subject matter and pedagogical knowledge) on student test performance at the district level. A more recent Texas study (Fuller, 1999) found that students of fully licensed teachers were significantly more likely to pass the Texas state achievement tests, after controlling for student socioeconomic status, school wealth, and teacher experience. In a recent school level analysis of mathematics test performance in California high schools, Fetler (1999) found a strong negative relationship between student scores and the

percentage of teachers on emergency certificates, as well as a smaller positive relationship between student scores and teacher experience levels, after controlling for student poverty rates.

Ferguson and Ladd's Alabama study. The Commission cited Ferguson and Ladd's (1996) study of student achievement in Alabama as additional support for the proposition that school inputs, particularly teachers (and to a lesser extent class sizes), matter for student learning. Ballou and Podgursky do not take issue with the findings as cited by the Commission. Their complaint about this study again concerns the Commission's definition of expertise, which included ACT scores, teacher experience, and master's degrees together. Here, too, the problem lies with Ballou and Podgursky's own narrow definition of the term, their misrepresentation of the Commission's proposals, and their desire to find malfeasance where none exists.

Greenwald, Hedges, and Laine's review. Ballou and Podgursky charge that the Commission's description of a meta-analysis of education production function studies by Greenwald, Hedges, and Laine (GHL) (1996) is "inaccurate." They once again misrepresent the Commission's actual statements when they claim: "The NCTAF describes this as a review of 'sixty production function studies,' and summarizes the GHL findings in a chart that purports to show the gain in student test scores per \$500 invested in various educational reforms.... The relevant table in GHL shows that the findings on teacher education are based not on 60 studies but eight." In fact, the meta-analysis did include 60 primary studies and the Commission's

summary of its major findings (below) is drawn directly from the study's conclusions and is entirely accurate. The Commission stated:

These findings about the influences and relative contributions of teacher training and experience levels are reinforced by those of a recent review of 60 production function studies (Greenwald, Hedges, and Laine, 1996) which found that teacher education, ability, and experience, along with small schools and lower teacher-pupil ratios, are associated with increases in student achievement across schools and districts. In this study's estimate of the achievement gains associated with expenditure increments, spending on teacher education swamped other variables as the most productive investment for schools (see figure 5). (Darling-Hammond, 1997, p. 9)

The chart produced in the Commission's report was reproduced directly from GHL's Table 6 (p. 378) with the same variable labels and statistics. Although GHL selected a subset of the studies for their analysis of the contribution of teacher education to student achievement, the table as a whole uses most or all of the 60 studies reviewed. There is nothing misleading about the Commission's chart. It directly represents the data as presented in the original source.

Ballou and Podgursky also call "misleading" the use of the variable label "teacher education" (the term used in the GHL table and replicated in the Commission's chart) for a variable that represents the salary investment associated with securing a master's

degree. This point is utterly mysterious. Do they mean to suggest that teacher learning on the job is not teacher education? They then suggest that the Commission indicated the student achievement gains would be based on a "one-time investment of \$500 in teacher training." This is patently false, as the legend on the chart makes clear. Finally, they object to a Commission claim that exists only in their own imagination. At no point did the Commission "claim that we can significantly raise student achievement by spending \$500 on teacher education." As the Commission's actual words, quoted above, make clear, the only claim is a report of the direct finding of the GHL study--that returns to investments in teacher education were found to be larger for a marginal increment of additional spending than they were found to be for other factors studied. While Ballou and Podgursky may not like the analysis or conclusions of the GHL study, the Commission did not misrepresent the study in any way.

Ballou and Podgursky also cite a critique of the GHL study by economist Eric Hanushek (1996) that questions what Hanushek called a "selective use of evidence" by GHL because they did not count all of the "studies" he suggested they should. Using verbal tactics that resemble those of Ballou and Podgursky, Hanushek (1996) accuses Greenwald, Hedges, and Laine of "manipulations," "systematic distortions," "systematic bias," and "misinterpretations" (pp. 397-398), and he presents new data that purport to counteract their analysis but offers no explanation of his criteria for selecting studies and coefficients. Ballou and Podgursky do not discuss the

response from Greenwald, Hedges, and Laine (1996b), which explains the reasons for these disparate findings:

We require that studies be independent, but Hanushek does not. Hanushek is content to count results based on the same individuals as many times as they happen to be reported, giving each weight equal to that of an independent replication. Thus if there are 10 coefficient estimates derived from the same individuals, Hanushek would count these as 10 "studies," but we would count them as 1 study. By counting results based on the same individuals more than once, Hanushek appears to have many more studies than we do.... Table 1 (in this rejoinder) shows that Hanushek generally counts results based on each data set more than once and that he counts the results that give negative (and negative statistically significant) results more often than those that give positive (and positive significant) results. By systematically overcounting negative results, Hanushek is able to achieve the appearance that the evidence is more evenly divided than we found it to be. (pp. 413-414)

Interested parties should read both analyses and draw their own conclusions. Attention to the actual data presented will perhaps allow the discourse to move beyond name calling to a real consideration of the evidence.

The Armour-Thomas study. In critiquing the Commission's reference to a study by Eleanor Armour-Thomas and colleagues

(1989) of demographically similar schools in New York City with very different student achievement outcomes, Ballou and Podgursky offer a circular argument that simultaneously suggests we should not be surprised that schools with high levels of teacher experience and education did substantially better and then argues there is no reason to believe these teacher attributes were critical causal factors. At one especially confused point in their discussion, they seem to be saying that, since teacher education levels and certification status are correlated with experience and stability (i.e., low turnover), none of them should be taken seriously as correlates of student learning. (In an endnote, Ballou and Podgursky also misinterpret the study methodology, assuming the study used individual teacher observations rather than average school ratings of teacher variables to measure teacher characteristics in the regression equations.)

There is some value to one of the points Ballou and Podgursky try to make regarding the potential importance of other schooling variables; however, they omit key data from their discussion that would have partially addressed the concern they raise. They indicate, properly, that measured teacher characteristics can, in fact, be a function of deeper causal factors not controlled for in the research. They note:

The leadership provided by school and district administrators and the degree of parental involvement have much to do with the success of the school. Much of this influence consists of intangibles or factors (e.g., school discipline and morale) rarely measured in

production function studies. As a result, the fact that higher-performing students are taught by teachers with more advanced degrees and experience can be a reflection of other factors not controlled for in the research, rather than a causal relationship.

While discounting the very strong relationship between teacher education, experience, and certification and student achievement in these schools, Ballou and Podgursky fail to note that the Armour-Thomas study also examined precisely the variables they list above—instructional and organizational leadership, instructional focus, teacher morale, and parental involvement. These "functional" school characteristics were correlated with teacher characteristics, which may mean that they do contribute to the development of a more stable and able staff or that they are partly a product of a more stable and able staff or that both co-vary with yet other unmeasured variables. However, these characteristics proved to be only very mildly related to student achievement, accounting for a small portion of the variance in average student scores (from 2% to 18% across equations assessing different grade levels and subject areas as compared to 86% to 96% for the teacher characteristics). Had the authors included all of these variables in the same regressions, it is likely that the functional characteristics would have absorbed some of the variance otherwise attributed to the teacher status characteristics. But it is virtually certain, given the strengths of the relationships, that the teacher characteristics would have remained highly significant. While this study, like others, does not answer all of

the questions one might wish it could, in combination with other research, it tells a consistent story.

NAEP data on correlates of reading achievement. At several junctures in their article, Ballou and Podgursky charge that the Commission "flatly misrepresents" research--a very serious charge. In reference to the Commission's citation of NAEP findings from the 1992 reading assessment regarding correlations between student reading scores and teacher qualifications and practices, Ballou and Podgursky charge that "the NAEP results in the commission's table differ from those reported by the U.S. Department of Education." They include a table that displays data they say they retrieved from a National Center for Education Statistics (NCES) website to illustrate alleged discrepancies. However, Ballou and Podgursky did not consult the NCES document which the Commission report actually used as its reference source (NCES, 1994). The two statistics they challenge are replicated in the Commission's report exactly as they appeared in this NCES document in Table 9.14 (pp. 360, 365). Ballou and Podgursky's charge that the "numbers the commission presented are simply wrong" is simply wrong. In addition, Ballou and Podgursky suggest that the Commission deliberately ignored data from the 1994 reading assessment that they allege contradicts the Commission's conclusions. However, the 1994 data they describe were not even available until 1996 when the Commission had completed this analysis and were not included in the source the Commission used.

Furthermore, and most ironically, the data Ballou and Podgursky report in their Table 2 do not match the statistics found at the

website address they cite

(nces.ed.gov/nationsreportcard/y25alm/almanac.shtml) in the website document that includes data from the NAEP 1992 and 1994 National Reading Assessments.(FN1) The data actually recorded on the website directly contradict Ballou and Podgursky's claims that "... in the 1994 results there is no difference between teachers with bachelor's degrees and those with master's degrees." In fact the website statistics show not only that students of teachers with master's degrees or higher continue to score higher than those of teachers with bachelor's degrees, but also that the weight of the data continues to buttress the Commission's major points about the value of teacher training. In their own selective use of data, Ballou and Podgursky fail to note that the findings of the 1994 survey (from the website they reference) show even larger differences than did the 1992 data in student achievement for teachers with substandard certification vs. regular or permanent certification.

Table 1 reproduces the statistics from the website. Item numbers and page numbers as recorded on the website are cited for those who would like to find the tables presenting these data. The 1994 data continue to show, as they did in 1992, higher scores for students whose teachers have had coursework in literature-based instruction along with strong support for teaching practices that use integrative approaches that combine work on reading and writing and use a range of texts. In addition, these data continue to report substantial differences--ranging from 10 to 20 points--in the average scores of students whose teachers used different teaching practices, which the NAEP study reports are strongly related to the teachers'

training. Use of a wide variety of trade books, newspapers, and other reading materials, frequent trips to the library, and integration of reading and writing continue to be associated with higher scores, while use of reading kits, workbooks, worksheets, and short-answer and multiple choice tests to assess reading continue to be associated with lower scores. The NAEP study found that the former practices are more likely to be used by teachers with more extensive training while the latter practices are more likely to be used by untrained teachers who rely on highly structured crutches to move students through independent seatwork.

Although the website statistics do not match those in their table, Ballou and Podgursky are correct that the 1994 data no longer show an advantage for students whose teachers have had course work in whole language instruction. The 1.6 point difference in average student scores (smaller than the gap they report) is smaller than the standard error of scores for the increasingly small percentage of teachers who have not had such course work--only 15% of the total in 1994 as compared to 20% in 1992. There is currently so little variability in this factor that differences in outcomes related to this variable should be increasingly difficult to measure. Ballou and Podgursky raise this point because they believe that whole language instruction, which they have caricatured as instruction "in which the teacher stands by while the student tries to guess what the word is" (Ballou & Podgursky, 1999b, p. 41) should be replaced by phonics instruction. In their own selective use of data, however, they fail to note in their review of the NAEP data that teachers' training in phonics-based instruction is associated with substantially lower

scores for students (see Table 1 and NCES, 1994).

The Commission did not introduce these data on phonics training in its earlier report both because the phonics vs. whole language debate has become so politically charged (with political conservatives including Ballou and Podgursky (1999b) arguing for phonics-based reading instruction in lieu of whole language approaches) and because it is not possible to know what kind of phonics-based training teachers may have been receiving. The teaching of reading should not be treated as an ideological question with one "side" trying to debunk the other. Research on the teaching of reading indicates that decoding skills should be taught within a larger framework of the many skills that comprise reading ability and as part of an approach that includes multiple language resources and experiences (National Research Council, 1998). Unfortunately, the teaching of phonics has too often assumed that knowledge of letter-sound relationships alone and out of context will enable students to read, an assumption that much research has shown to be flawed. It has sometimes led to a particularly unsophisticated approach to reading instruction that often focuses almost entirely on fill-in-the-blank practice with discrete subskills and letter-sound comparisons (managed through the reading kits, workbooks, and worksheets also found to be associated with lower reading scores) without integration of these skills into authentic reading and writing activities that develop students' abilities to use syntax and context clues, interpret what they read, and develop comprehension (Cooper & Sherk, 1989). As economists like Ballou and Podgursky go outside of their field of expertise to try to make claims about pedagogical

strategies, they, too, should be held to standards of proper use of research and evidence.

I do not know why the statistics Ballou and Podgursky cite do not match those in the NCES print document on 1992 reading assessment background data or the NCES website document on the 1992 and 1994 reading scores and teacher characteristics. I could not find the 8th grade data they cite in my visit to the NCES website. However, assuming there is a legitimate source for their data, there is no point debating which of the sources of these NCES statistics is more "right" than the other. There are frequently differences between statistics cited in different data sources for reasons ranging from typographical errors to reestimations of statistics due to data cleaning, recoding, collapsing of categories, decisions about how to handle missing data, and the like. However, charges of deliberate misrepresentation of data are very serious. Making such charges without ascertainment of sources and accurate rendering of claims may be acceptable in the political realm, but it violates the ethical norms of the research community.

STATE NAEP TRENDS

Ballou and Podgursky also charge selective use of evidence in the Commission's brief summary of an analysis of state NAEP score trends published in *Doing What Matters Most* (DWMM; Darling-Hammond, 1997). They complain that only 11 of 36 potential cases were displayed for 4th grade math and 9 of 37 in reading (actually there were 13 of 36 cases displayed for 4th grade math and 15 for 8th grade math) and that the analysis excluded Texas, a state with high reported gain scores. The discussion in DWMM was based on a

50 state survey of state policies conducted by the Commission and augmented with data from state policy surveys conducted by the Council for Chief State School Officers and the American Association of Colleges for Teacher Education and from state case studies. Aside from the obvious fact that it would be impossible to read a chart that displayed 36 different trend lines, the selection of states for display and discussion was a function of three factors:

- * States that had enacted extraordinarily comprehensive reforms of teaching during the late 1980s and 1990s (Connecticut and North Carolina enacted the most ambitious teacher legislation of any states nationally, followed by Arkansas, Kentucky, and West Virginia)

- * States that enacted the earliest and most comprehensive test-based accountability laws in the early 1980s (Georgia and South Carolina)

- * States that participated in both testing years and that scored at the very top or bottom of the achievement distribution in the years displayed, illustrating the range of scores (Connecticut, Iowa, Maine, Minnesota, Montana, North Dakota, Nebraska, and Wisconsin at the top and Louisiana and Mississippi at the bottom)

The analysis examined trends in achievement in the states with comprehensive, long-standing teacher investments and in the states with comprehensive, long-standing test-based accountability policies (states which also happened to have made little investment in teacher policy reform or professional development during the 1980s and '90s), addressing the logical and timely policy question of what outcomes might be associated with these different reform strategies.

In addition, the analysis examined the teaching policies of the top-scoring states. The remarkable similarities in their policies and strategies led to another analysis of state teaching policies and student achievement (Darling-Hammond, 1999) that is summarized below.

Ballou and Podgursky question why Texas was not among the states discussed in this analysis, noting that it registered math gains as great as North Carolina in 4th grade mathematics and substantial in 8th grade mathematics and suggesting that the Commission chose only the evidence that agreed with its thesis. (They do not note that Texas registered declines in reading achievement during the years measured.) In addition to the fact that Texas did not meet the original criteria for selection into the analysis, noted above, there are many reasons to doubt the stability of scores in Texas and the extent to which the posted gains are real. As noted in an endnote in the report (Darling-Hammond, 1997, p. 45), Texas included fewer than 45% of its students with disabilities in the testing pool, according to NAEP's original inclusion criteria (NCES, 1997, Table D3). Excessive exclusions of low-scoring students from the testing pool can cause gain scores to appear much larger than they would otherwise be. In addition, recent studies in Texas have raised concerns that much of the ostensible gain registered by African American and Latino students has been a function of widespread grade retentions and dropouts/pushouts that have increased substantially in recent years. These practices also make average test scores look higher by eliminating lower scoring students from the testing pool and, often, from school altogether (Kurtz, 1999;

Mexican American Legal Defense and Education Fund, 1999). The state is currently being sued for the discriminatory impact on Black and Latino students of its testing policies.

What if some portion of the gains in Texas is real? What policies might be responsible? In a report to the National Education Goals Panel, Grissmer and Flanagan (1998) highlighted Texas as one of the states posting large gains and attributed these gains primarily to the state's accountability system. They also mention the state's shifts of resources to more disadvantaged students through school finance equalization, class size reductions, and the creation of full day kindergarten. The school funding investments that occurred in the 1980s and were continued into the following decade may indeed have made some difference in Texas students' achievement in the 1990s. However, the state's new assessment and accountability system was not initiated until 1994 and not fully implemented until 1995-1996, so it could not have accounted for gains between 1990 and 1996. In addition to the equalization of funding and investments in reduced class sizes and kindergarten pointed out above, Texas was among the few states recognized by the National Education Goals Panel (1998) for large gains in the proportion of beginning teachers receiving mentoring from expert veterans. Texas has also had a growing number of 5-year teacher education programs in response to an earlier reform eliminating teacher education majors at the undergraduate level. There may be policy lessons to learn from Texas, but until the concerns about high rates of test exclusion, grade retention, and dropouts are resolved, it is difficult to know what they are.

BALLOU AND PODGURSKY'S "PROFESSIONALIZATION SCALE"

Perhaps the most amazing part of Ballou and Podgursky's critique rests on a strange analysis they conducted to correlate "ratings" they invented from a policy checklist in the appendix of the Commission's report with average state student achievement scores. Much of their criticism of the Commission's analysis of state achievement trends (above) refers to state "ratings" on what they call a "ten-point scale." They argue that the Commission's recommendations must be flawed because some states with high levels of student achievement or achievement gains have relatively low "ratings" on this so-called "professionalization scale." What Ballou and Podgursky are referring to is a set of indicators included in the report's appendix and labeled "Indicators of Attention to Teaching Quality" that were assembled from nationally available data regarding both teachers and state policies.

In the first place, the indicators were not developed as a scale or as a measure of professionalization. They refer to a number of the statistics cited in the Commission report on issues ranging from teacher preparation, licensing, and certification to proportions of teachers versus other school staff and policies influencing incentives for such things as National Board Certification. They were chosen for both their interest to Commissioners and members of the public and their ready availability on a cross-state basis. The purpose of the appendix (one of many that provided state-by-state data) was to provide a partial answer to the frequent questions, "What's going on in our state?" and "How do states vary?" on a set of indicators of

general interest.

The idea of correlating a count of these indicators with student achievement scores is a bizarre idea for several reasons: First, the indicators were not designed to comprise any sort of scale and make no sense when viewed in that way. They are certainly no measure of "professionalization," if such could be devised. Second, many of the indicators would not be expected to influence large-scale student achievement directly in any case (for example, the proportion of teachers as a percentage of all staff, which was offered as an indicator of resource allocation decisions). Third, many of the policy indicators were just enacted by the states in 1995 or 1996 and had not even been implemented yet (e.g., National Board incentives, several states' professional standards boards, many new teacher induction programs). They could not be expected to influence student achievement for many years, and would certainly not affect achievement trends that occurred before they were even enacted! Given the indefensibility of the Ballou and Podgursky analysis, it is surprising that they derived positive coefficients for all of their regressions of the so-called scale scores on student achievement.

A MORE VALID MEASURE OF THE EFFECTS OF TEACHING STANDARDS

A more valid way to gauge the influences of better qualified teachers on student learning is to measure the relationship directly. We did exactly that in an analysis using several of the indicators of teacher qualifications reported in the appendix of the Commission's 1997 report. The results, reported more fully elsewhere (Darling-Hammond, 1999), provide strong support for the thesis that well-

qualified teachers make a substantial difference in student learning, and that specific policies may heighten the proportion of well-qualified teachers in a state.

After controlling for student poverty and for student language background (LEP status) in reading, the most consistent and highly significant predictor of average student achievement in reading and mathematics on NAEP tests from 1990 through 1996 was the proportion of well-qualified teachers in a state: those with full certification and a major in the field they teach. In regression estimates that took into account class sizes, teachers' education levels and certification status, along with student characteristics, the teacher quality variables accounted for between 40% and 60% of the total variance in student achievement in reading and mathematics in each of the years and grade levels assessed.

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 (Byrne, 1983; Darling-Hammond, 1999), it makes sense that this variable would be more powerful than measures of either subject matter knowledge or teaching knowledge alone. It is also possible that this variable captures other features of the state policy environment, including general investments in education as well as aspects of the regulatory system for education, such as the extent to which standards are rigorous and the extent to which they are enforced, issues to which we return later.

RESEARCH FINDINGS ON TEACHER EDUCATION AND EXPERTISE

This is the nub of the matter and the central point on which the Commission and Mrs. Ballou and Podgursky disagree. Ballou and Podgursky object to the Commission's recommendations that would improve and extend preservice teacher education, questioning whether there is any teacher input beyond general literacy or basic academic ability that is related to teaching effectiveness.

Focusing in on only four studies among the 92 references cited in the Commission's two reports that included findings concerned with teacher education, professional development, and teaching effects (many of them research reviews that incorporate dozens of studies), Ballou and Podgursky suggest that "there is no evidence in these studies that teachers should be 'better trained' in any specific sense.... The studies cited above offer no support for any particular pedagogical approach over another." They repeatedly reiterate their criticism that the four studies they have chosen to focus on do not prove the importance of teacher education.

These disingenuous statements ignore the fact that the four studies they have selected were cited for the general proposition that what teachers know and can do matters, while the Commission cited dozens of other studies on issues of teacher education, including many that describe features of teaching and teaching knowledge that have been found particularly important for teaching success as evaluated through supervisory ratings, researchers' observations of teaching practices, and student learning gains. Commission reports that are to be read by large general audiences are not the place for

detailed reviews of the literature on teaching effectiveness. However, many relevant reviews and individual studies are referenced in the endnotes and reviewed more extensively in other publications. (For a recent review, see Darling-Hammond, 1999.)

Elsewhere, in a rambling attack on constructivist learning theory and modern theories of intelligence that shows their misunderstanding of these and other concepts, Ballou and Podgursky (1999b) argue against requirements for preservice education or even for specific subject matter training for teachers, suggesting that subject matter tests are a better indicator of competence. They argue that pedagogical knowledge is unimportant. Many of the claims they make cannot be supported by the research literature that is available--research of which they do not appear to be aware. Much of what the literature suggests about the relative importance of general academic ability, and pedagogical and subject matter knowledge contradicts Ballou and Podgursky's (1999b) unsupported claims in the Fordham publication and their critique of the Commission's report that policies should aim to recruit individuals with high levels of "general intelligence and academic ability" while minimizing exposure to knowledge about teaching or extensions of training that they claim would deflect capable people from the profession.

Their argument ignores evidence that measures of general intelligence have not been found to correlate strongly with teaching effectiveness (Schalock, 1979; Soar, Medley, & Coker, 1983); that tests of subject matter knowledge have shown little relationship to teaching performance, while measures of subject matter preparation

that examine course taking have shown a stronger relationship to teaching performance (Andrews, Blackmon, & Mackey, 1980; Ayers & Qualls, 1979; Haney, Madaus, & Kreitzer, 1986; Hawk, Coble, & Swanson, 1985; Monk, 1994; Quirk, Witten, & Weinberg, 1973); and that knowledge about teaching and learning shows even stronger relationships to teaching effectiveness than subject matter knowledge (Ashton & Crocker, 1987; Begle, 1979; Begle & Geeslin, 1972; Denton & Lacina, 1984; Druva & Anderson, 1983; Evertson, Hawley, & Zlotnik, 1985; Ferguson & Womack, 1993; Guyton & Farokhi, 1987; Monk, 1994; Perkes, 1967-1968).

Their argument also ignores evidence that better-prepared teachers, such as those recruited through 5-year teacher education programs, are not only more academically able than those recruited through 4-year programs, they actually enter and remain in teaching at higher rates and are found to be more effective (Andrew, 1990; Andrew & Schwab, 1995; Arch, 1989; Denton & Peters, 1988). With most current college students taking at least 5 years to complete a college degree anyway and those entering teaching wanting to be adequately prepared, it is not at all clear that providing year-long internships for teaching connected to more coherent coursework on content and teaching would be a disincentive to enter the field.

Furthermore, those who enter with little professional preparation--a practice Ballou and Podgursky (1997a, 1999b) endorse--tend to have greater difficulties in the classroom (Bents & Bents, 1990; Darling-Hammond, 1992; Lenk, 1989; Feiman-Nemser & Parker, 1990; Gomez & Grobe, 1990; Grady, Collins, & Grady, 1991; Grossman, 1989; Mitchell, 1987; National Center for Research on

Teacher Learning, 1992; Rottenberg & Berliner, 1990), are less highly rated by principals, supervisors, and colleagues (Bents & Bents, 1990; Jelmberg, 1995; Lenk, 1989; Feiman-Nemser & Parker, 1990; Gomez & Grobe, 1990; Mitchell, 1987; Texas Education Agency, 1993), and tend to leave teaching at higher-than-average rates, often reaching 50 percent or more by the third year of teaching (Darling-Hammond, 1992; Lutz & Hutton, 1989; Stoddart, 1992).

Finally, Ballou and Podgursky's argument ignores the historical evidence from other professions that more rigorous, sometimes lengthened preparation, has increased the general ability and the professional knowledge of entrants as heightened standards have improved salaries and working conditions at the same time. (The Commission report also illustrates how increased investments in teachers can be supported by reallocating funds within current education budgets that invest much less in teachers in the United States than do other countries.)

We agree with Ballou and Podgursky that the "connection between reforms and research needs to be much more carefully drawn." We wish that Ballou and Podgursky would read the research they ignore that is cited in the documents they attack and that they would familiarize themselves with the bodies of additional research that bear upon the subjects upon which they make pronouncements.

THE ARGUMENT FOR PROFESSIONAL TEACHING STANDARDS

Ballou and Podgursky disagree with the need for licensing standards that would require knowledge of teaching and learning and evidence of teaching ability as well as subject matter knowledge

for teachers; they take issue with professional accreditation as a vehicle for ensuring quality control for teacher education; and they dismiss the National Board for Professional Teaching Standards as the work of "moonlighting" public school teachers without having examined the Board's rigorous tests, which are developed and scored with the help of the Educational Testing Service (ETS) and are buttressed by 10 years of research establishing their reliability and validity. These three sets of standards are aligned with one another and with new standards for student learning in the disciplines, they are based on current research on teaching and learning, and they are tied to performance-based assessments of teacher knowledge and skill. The standards see teaching as responsive to student learning, rather than the mere implementation of routines. The assessments look at evidence of teaching ability (videotapes of teaching, lesson plans, student work, analyses of curriculum) in the context of real teaching and in relation to evidence of student learning.

At the heart of the Commission's recommendations regarding standards are the ideas that professional standards are a lever for raising the quality of practice and that they are central to the cause of equity, protecting especially the least advantaged clientele from incompetent practitioners. Having a floor of competence below which no practitioner will be allowed to fall is especially critical in cases where there is unequal spending ability coupled with lack of access to information about competence. Whereas high-spending districts may be able to engage in expensive searches and hiring tests and to purchase well-qualified candidates regardless of what external

standards exist, poorer districts are less likely to be able either to evaluate candidates without external information about their competence or to recruit adequately qualified candidates without requirements that safeguard the well-being of the students they serve. Without standards, there is no basis upon which to hold states accountable for funding and overseeing districts in ways that protect students' rights to adequately qualified teachers. Without standards, there are also few means by which to ensure the spread of professional knowledge.

Standard-setting is at the heart of every profession. When people seek help from doctors, lawyers, accountants, engineers, or architects, they rely on the unseen work of a three-legged stool supporting professional competence: accreditation, licensing, and certification. In virtually all professions other than teaching, candidates must graduate from an accredited professional school that provides up-to-date knowledge in order to sit for the state licensing examinations that test their knowledge and skill. The accreditation process is meant to ensure that all preparation programs provide a reasonably common body of knowledge and structured training experiences that are comprehensive and current.

Licensing examinations, developed by members of the profession through state professional standards boards, are meant to ensure that candidates have acquired the knowledge they need to practice responsibly. States do not abandon their responsibility for licensure when they establish such boards in each of the professions. They delegate the responsibility for grappling with the nuances of standard-setting to members of the profession and members of the

public who are publicly appointed by governors or state boards of education. Public accountability continues to exist through the appointment process and the fact that such boards must secure board or legislative approval for many of their policies and expenditures of funds. The tests generally include both surveys of specialized information and performance components that examine aspects of applied practice in the field: Lawyers must analyze cases and develop briefs or memoranda of law to address specific issues; doctors must diagnose patients via case histories and describe the treatments they would prescribe; engineers must demonstrate that they can apply certain principles to particular design situations.

In addition, many professions develop standards and examinations that provide recognition for advanced levels of skill, such as certification for public accountants, board certification for doctors, and registration for architects. The certification standards are used not only to designate higher levels of competence but also to ensure that professional schools incorporate new knowledge into their courses and that practitioners incorporate such knowledge in their practice. They also guide professional development and evaluation throughout the career. Thus, these advanced standards may be viewed as the engine that pulls along the knowledge base of the profession. Together, standards for accreditation, licensing, and certification support quality assurance and help ensure that new knowledge will be incorporated into training and used in practice.

These are the major policy levers available for influencing profession-wide learning. It is for this reason that the Commission focused some of its attention on reforms of these functions and the

creation of professional standards boards that administer them. Ballou and Podgursky may not understand how accreditation, licensure, and advanced certification are used in other professions to ensure quality control and advance the spread of knowledge. Or they may sympathize with those interested in maintaining low cost teacher education (no teacher education at all would be the least expensive) and low standards for teaching. These conditions support low salaries and lack of investment in teachers' knowledge, and allow great inequalities in the teaching skills to which most rich and poor students are exposed.

While Ballou and Podgursky object to standards boards, licensing standards, and National Board certification, they have spent most of their energy arguing against accreditation of education schools (Ballou & Podgursky, 1997a, 1997b, 1999b). They have based their argument in part on the claim that colleges that have higher general admission standards for their students are less likely to be NCATE-accredited and in part on the claim that their studies do not show higher outcomes for NCATE graduates. While the Commission's recommendation regarding professional accreditation of education schools is by no means its only or most important recommendation, the charges they make deserve a response.

First, Ballou and Podgursky's claims regarding the relative selectivity of state colleges that are often NCATE accredited and private institutions that often are not rely on the wrong indicators of candidate qualifications. Because of state reforms in the 1980s, most state institutions and many private ones have established higher standards for admission to their teacher education programs

than they maintain for general college admission. Consequently, it is not appropriate to assume that general college admissions standards are comparable to those of teacher education programs. An analysis of the federal Baccalaureate and Beyond database found that 1993 graduates of NCATE-accredited teacher education programs were 50% more likely to have scored above the 50th percentile on SAT and ACT tests than graduates of non-NCATE teacher education programs (Shotel, 1998). NCATE graduates had also taken 25% more social science, computer science, advanced foreign language credit, precollege mathematics, and teaching coursework and 25% fewer remedial English and statistics courses than non-NCATE graduates, with other areas being approximately equal (Shotel, 1998). A more recent analysis by ETS found that among 270,000 test-takers in 1995 through 1997, graduates of NCATE-accredited colleges of education passed ETS's content examinations for teacher licensing (the Praxis subject matter tests) at a significantly higher rate than did graduates of unaccredited programs, boosting their chances of passing the examination by nearly 10 percent (Gitomer, Latham, & Zimek, 1999). In the teacher quality study mentioned earlier (Darling-Hammond, 1999), the strongest predictor of a state's percentage of well-qualified teachers (that is, teachers with both a major and full certification in their field) was the percentage of teacher education institutions in a state that meet national accreditation standards through NCATE ($r = .42$, $p < .05$).

This is not to argue that NCATE accreditation is the perfect measure of institutional quality. Because teacher education

accreditation is voluntary in most states, some high-quality programs that are unaccredited would have little difficulty meeting the NCATE standards, but have had no incentive for pursuing accreditation. Many others would not meet the standards because the content, coherence, and resources of their programs are inadequate. Studies indicating that negative NCATE reviews have led to substantial changes and investments in weak education programs (see, for example, Altenbaugh & Underwood, 1990) also highlight the fact that professional accreditation can be at odds with universities' desires to use education schools as "cash cows" for other parts of the university. (See Howard, Hitz, and Baker, in press, for evidence that schools of education continue to be less well funded than other professional programs.)

Why so much attention to the issue of accreditation? Gallagher and Bailey (in press) describe how the professionalization of medicine depended on efforts of the American Medical Association's Council on Medical Education in supporting Abraham Flexner's research, establishing a system of accreditation, and encouraging states not to grant licenses to graduates of poorly rated schools. The American Bar Association undertook similar efforts to upgrade the quality of law schools in the early part of the century. In these and other professions, the insistence on accreditation of training institutions was a primary means of raising standards, infusing common knowledge into professional schools, and ensuring that all entrants got access to this knowledge.

Transforming preparation in a profession as a whole is no small task. As one indication of the gap between program practices and

professional standards, the initial failure rate for programs seeking accreditation in the three years after NCATE strengthened its standards in 1987 was 27%. During the first three years of implementation, almost half of the schools reviewed could not pass the new "knowledge base" standard, which specified that schools must be able to describe the knowledge base on which their programs rest. Most of these schools made major changes in their programs since that time, garnering new resources, making personnel changes, and revamping curriculum, and were successful in their second attempt at accreditation.

More than 90% of all programs that have stood for accreditation in the last several years report that the process led to major improvements in the quality of their programs. NCATE upgraded its standards again in 1995 to incorporate the Interstate New Teacher Assessment and Support Consortium (INTASC) and National Board standards, and it has even more ambitious plans for performance-based accreditation by the year 2000. This means that many programs that want to secure or maintain professional accreditation will need to upgrade their efforts further. There are unresolved financial, political, and substantive issues that will determine how many programs undertake these efforts.

There is more at stake here than governmental and institutional turf battles. As the move for professional standards has become stronger, as licensing, certification, and accreditation standards have become more rigorous and more performance-based, the continuation of teacher education as a low-status, low-expenditure operation that provides revenues for other parts of the university

could be threatened. So, too, could the continuation of teaching as an enterprise in which low salaries are maintained by easy access and low standards. Finally, the continuation of unequal resource allocations supported by lack of state responsibility to adhere to standards of care for all students might unravel if standards of teacher preparation and licensing were as strong as those in other professions.

Is this a top-down regulatory approach at odds with decentralized performance-driven reforms, as Ballou and Podgursky (1997b) have charged elsewhere? We believe it is not. Far from usurping the authority of parents and community members, standards for teaching can help ensure that all students have teachers who are better equipped to work with and support students and families. This could ultimately promote many different models of education, because many regulations that currently constrain schools will be unnecessary if the state takes care of its key obligation: the preparation and equitable distribution of highly qualified teachers who know how to do their jobs well. If policy makers and the public are convinced that educators are well-prepared to make sound decisions, they should find it less necessary to regulate schools against the prospect of incompetence. Assuring quality in the teaching force is actually, we believe, the best way to support decentralization and local control in education.

CONCLUSION

Ballou and Podgursky have ignored or misrepresented most of the existing evidence base in order to argue that teacher education makes no difference to teacher performance or student learning and

that students would be better off without state efforts to regulate entry into teaching or to provide supports for teachers' learning. While they argue for recruiting bright people into teaching (and who could disagree with that?), their proposals offer no incentives for attracting individuals into teaching other than the removal of preparation requirements. While they present this as an attraction to teaching, evidence suggests that lack of preparation actually contributes to high attrition rates and thereby becomes a disincentive to long-term teaching commitments and to the creation of a stable, high ability teaching force. Lack of preparation also contributes to lower levels of learning, especially for those students who most need skillful teaching in order to succeed.

We agree with one of Ballou and Podgursky's concluding statements: "Unlike markets for medical or other professional services, most education consumers (parents and children) have little choice as to their teachers or schools. Given that it is a captive market, the potential for harm from such a policy cannot be ignored." The evidence we have presented here and elsewhere makes clear that the policies Ballou and Podgursky endorse would bring harm to many children, especially those who are already least well served by the current system. They should bear the burden of proof for showing how what they propose could lead to greater equity and excellence in American schools.

ADDED MATERIAL

LINDA DARLING-HAMMOND is Charles E. Ducommun Professor of Education at Stanford University and Executive Director of the National Commission on Teaching and America's Future, which

produced the report *What Matters Most: Teaching for America's Future*. Her research focuses on teaching, school organizations and reform, and educational equity. She is author of *The Right to Learn* (Jossey-Bass, 1997).

Table 1. Average student proficiency scores in reading (NAEP, 1994) by selected teacher characteristics and practices

Correlates of reading achievement	Lower scores	
Higher scores	Teaching certification	Temporary
(substandard) Highest level T050001, p. 31	31	204.9
(5.0) 214.5 (1.6) Certification in reading	No	
Yes T040502, p. 46	212.8 (1.4)	216.4 (2.8)
Degree level	Bachelor's degree	Master's
degree T050101, p. 81	212.5 (1.3)	215.01
(1.9) Course work in literature-instruction	No	Yes based
210.2 (2.9)	214.2 (1.2)	T045907, p. 271
Course work in whole instruction	Yes	No language
213.4 (1.2)	215.0 (3.5)	T045911, p. 291
Course work in phonics	Yes	No
T045912, p. 296	205.8 (1.9)	216.7 (1.5)
Types of materials used	Basal readers	Trade books
T046501, p. 356	207.9 (2.4)	217.8 (2.9)
Frequency with which teachers use reading kits	Every day	Never
199.3 (4.2)		215.9 (1.5)
T046602, p. 371	Frequency with which students	Every day
Never use workbooks or worksheets		207.1 (2.5)

219.1 (2.7) T046705, p. 451 Frequency with which teachers
 Never At least once a week take class to library

202.1 (7.3) 214.2 (1.1) T047101, p. 541 Frequency with
 which students Never Every day write about what
 they have read 209.9 (7.0) 212.4 (2.5) T046704, p. 441

Use of multiple choice tests At least once a week Once or
 twice/year to assess students in reading 210.7 (3.1)

215.5 (3.2) T047001, p. 511 Use of short-answer tests to At
 least once a week Once or twice/year assess students in reading

208.6 (1.9) 219.7 (4.4) Use of written assignments to
 Never At least once a week assess students in reading

200.8 (3.8) 215.2 (1.5) T047003, p. 521

FOOTNOTE

1 The website document is titled "NAEP 1992, 1994 National Reading Assessments--Data Almanac--Grade 4: Teacher Questionnaire Weighted Percentages and Composite Proficiency Means (Public Schools)."

REFERENCES

- Altenbaugh, R. J. & Underwood, K. (1990). The evolution of normal schools. In J. I. Goodlad, R. Soder, & K. Sirotnik (Eds.), *Places where teachers are taught* (pp. 136-186). San Francisco: Jossey-Bass.
- Andrew, M. (1990). The differences between graduates of four-year and five-year teacher preparation programs. *Journal of Teacher Education*, 41(2): 45-51.
- Andrew, M., & Schwab, R. L. (1995). Has reform in teacher education influenced teacher performance? An outcome assessment

of graduates of eleven teacher education programs. *Action in Teacher Education*, 17(3): 43-53.

Andrews, J. W., Blackmon, C. R., & Mackey, J. A. (1980). Preservice performance and the national teacher examinations. *Phi Delta Kappan*, 61(5): 358-359.

Arch, E. C. (1989). Comparison of student attainment of teaching competence in traditional preservice and fifth-year master of arts in teaching programs. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Armour-Thomas, E., Clay, C., Domanico, R., Bruno, K., & Allen, B. (1989). An outlier study of elementary and middle schools in New York City: Final report. New York: New York City Board of Education.

Ashton, P., & Crocker, L. (1987). Systematic study of planned variations: The essential focus of teacher education reform. *Journal of Teacher Education*, 38(3): 2-8.

Ayers, J. B., & Qualls, G. S. (Nov./Dec. 1979). Concurrent and predictive validity of the national teacher examinations. *Journal of Educational Research*, 73(2): 86-92.

Ballou, D., & Podgursky, M. (1997a). Reforming teacher training and recruitment: A critical appraisal of the recommendations of the National Commission on Teaching and America's Future. *Government Union Review*, 17(4): www.psr.org/doc/v174_art.html.

Ballou, D., & Podgursky, M. (1997b). Reforming teacher training and recruitment: A critical appraisal of the recommendations of the National Commission on Teaching and America's Future. *Opportunity*, 2(1): 4-5, 11-13.

Ballou, D., & Podgursky, M. (1999a). Reforming teacher preparation and licensing: What is the evidence? Teachers College Record <http://www.tcrecord.org> ID: 10418 (10/16/99).

Ballou, D., & Podgursky, M. (1999b). Teacher training and licensure: A layman's guide. In M. Kanstoroom & C. E. Finn (Eds.), *Better teachers, better schools*. Washington, DC: The Fordham Foundation.

Begle, E. G. (1979). *Critical variables in mathematics education*. Washington, DC: Mathematical Association of America and National Council of Teachers of Mathematics.

Begle, E. G., & Geeslin, W. (1972). *Teacher effectiveness in mathematics instruction*. National Longitudinal Study of Mathematical Abilities Report No. 28. Washington, DC: Mathematical Association of America and National Council of Teachers of Mathematics.

Bents, M., & Bents, R. (1990). Perceptions of good teaching among novice, advanced beginner and expert teachers. Paper presented at the Annual Meeting of the American Educational Research Association, Boston, MA.

Berliner, D. C. (1986). In pursuit of the expert pedagogue, *Educational Researcher*, 15(6): 5-13.

Byrne, C. J. (1983). Teacher knowledge and teacher effectiveness: A literature review, theoretical analysis and discussion of research strategy. Paper presented at the meeting of the Northwestern Educational Research Association, Ellenville, NY.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). *Equality of*

educational opportunity. Washington, DC: U.S. Government Printing Office.

Cooper, E., & Sherk, J. (1989). Addressing urban school reform: Issues and alliances. *Journal of Negro Education*, 58(3): 315-331.

Darling-Hammond, L. (1992). Teaching and knowledge: Policy issues posed by alternative certification for teachers. *Peabody Journal of Education*, 67(3): 123-154.

Darling-Hammond, L. (1997). *Doing what matters most: Investing in quality teaching*. New York: National Commission on Teaching and America's Future.

Darling-Hammond, L. (1998). Teachers and teaching: Testing policy hypotheses from a national commission report. *Educational Researcher*, 27(1): 5-15.

Darling-Hammond, L. (in press, 1999). *Teacher quality and student achievement: A review of state policy evidence*. Seattle: Center for the Study of Teaching and Policy, University of Washington.

Darling-Hammond, L., Hudson, L., & Kirby, S. (1989). *Redesigning teacher education: Opening the door for new recruits to science and mathematics teaching*. Santa Monica: The RAND Corporation.

Denton, J. J., & Lacina, L. J. (1984). Quantity of professional education coursework linked with process measures of student teaching. *Teacher Education and Practice*, 1(1): 39-64.

Denton, J. J., & Peters, W. H. (1988). *Program assessment report: Curriculum evaluation of a non-traditional program for certifying teachers*. Texas A&M University, College Station, TX.

Druva, C. A., & Anderson, R. D. (1983). Science teacher characteristics by teacher behavior and by student outcome: A meta-analysis of research. *Journal of Research in Science Teaching*, 20(5): 467-479.

Evertson, C., Hawley, W., & Zlotnick, M. (1985). Making a difference in educational quality through teacher education. *Journal of Teacher Education*, 36(3): 2-12.

Feiman-Nemser, S., & Parker, M. B. (1990). Making subject matter part of the conversation or helping beginning teachers learn to teach. East Lansing, MI: National Center for Research on Teacher Education.

Ferguson, P., & Womack, S. T. (1993). The impact of subject matter and education coursework on teaching performance. *Journal of Teacher Education*, 44(1): 55-63.

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 Helen Ladd (Ed.) *Holding schools accountable* (pp. 265-298). Washington, DC: Brookings Institution.

Fetler, M. (1999). High school staff characteristics and mathematics test results. *Education Policy Analysis Archives*, 7 (March 24), <http://epaa.asu.edu>.

Finn, C. E. (1999). Preface. In M. Kanstoroom & C. E. Finn (Eds.), *Better teachers, better schools* (pp. v-vii) Washington, DC: The Fordham Foundation.

Fuller, E. J. (1999). Does teacher certification matter? A comparison of TAAS performance in 1997 between schools with low and high percentages of certified teachers. Austin: Charles A. Dana Center, University of Texas at Austin.

Gallagher, K., & Bailey, J. (in press). Strategic philanthropy. In J. Bailey & K. Gallagher, *Politics of education yearbook*. Corwin Press.

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.

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.

Grady, M. P., Collins, P., & Grady, E. L. (1991). Teach for America 1991 Summer Institute evaluation report. Unpublished manuscript.

Greenwald, R., Hedges, L. V., & Laine, R. D. (1996). The effect of school resources on student achievement. *Review of Educational Research*, 66(3): 361-396.

Greenwald, R., Hedges, L. V., & Laine, R. D. (1996). Interpreting research on school resources and student achievement, *Review of Educational Research*, 66(3): 411-416.

Grissmer, D., & Flanagan, A. (1998). Exploring rapid achievement gains in North Carolina and Texas. Washington, DC: National Education Goals Panel.

Grossman, P. L. (1989). Learning to teach without teacher education, *Teachers College Record*, 91(2): 191-208.

Guyton, E., & Farokhi, E. (1987). Relationships among academic performance, basic skills, subject matter knowledge and teaching skills of teacher education graduates. *Journal of Teacher Education*, 38(5): 37-42.

Haney, W., Madaus, G., & Kreitzer, A. (1987). Charms talismanic: Testing teachers for the improvement of American education. In E. Z. Rothkopf (Ed.) *Review of research in education* (Vol. 14, pp. 169-238). Washington, DC: American Educational Research Association.

Hanushek, E. (1996). A more complete picture of human resource policies. *Review of Educational Research*, 66(3): 397-409.

Hawk, P., Coble, C. R., & Swanson, M. (1985). Certification: It does matter. *Journal of Teacher Education*, 36(3): 13-15.

Howard, R. D., Hitz, R., & Baker, L. (in press). Are education schools cash cows or money pits? In J. Bailey & K. Gallagher, *Politics of education yearbook*. Corwin Press.

Jelmborg, J. R. (1996). College based teacher education versus state sponsored alternative programs. *Journal of Teacher Education*, 47(1): 60-66.

Kurtz, M. (1999). Schools, teens battle barrier of 9th grade. *American-Statesman* (May 23).

Lenk, H. A. (1989). A case study: The induction of two alternate route social studies teachers. Unpublished doctoral dissertation. Teachers College, Columbia University.

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.

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 matter 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.

National Center for Education Statistics. (1994). *Data compendium for the NAEP 1992 reading assessment of the nation and the states: 1992 NAEP trial state assessment*. Washington, DC: U.S. Department of Education.

National Center for Research on Teacher Learning (NCRTL). (1992). *Findings on learning to teach*. East Lansing, MI: National Center for Research on Teacher Learning.

National Commission on Teaching and America's Future (NCTAF). (1996). *What matters most: Teaching for America's future*. New York: Author.

Perkes, V. A. (1967-1968). Junior high school science teacher preparation, teaching behavior, and student achievement. *Journal of Research in Science Teaching*, 6(4): 121-126.

Quirk, T. J., Witten, B. J., & Weinberg, S. F. (1973). Review of studies of concurrent and predictive validity of the national teacher examinations. *Review of Educational Research*, 43(1): 89-114.

Rottenberg, C. J., & Berliner, D. C. (1990). Expert and novice teachers' conceptions of common classroom activities. Paper presented at the Annual Meeting of the American Educational Research Association, Boston, MA.

Schalock, D. (1979). Research on teacher selection. In D. C. Berliner (Ed.), *Review of research in education* (vol. 7). Washington, DC: American Educational Research Association.

Shotel, J. (1998). Analysis of the 1993 Baccalaureate and Beyond longitudinal study. Unpublished manuscript, George Washington University, Washington, DC.

Skipper, C. E., & Quantz, R. (1987). Changes in educational attitudes of education and arts and science students during four years of college. *Journal of Teacher Education*, 38(3): 39-44.

Soar, R. S., Medley, D. M., & Coker, H. (1983). Teacher evaluation: A critique of currently used methods. *Phi Delta Kappan*, 65(4): 239-246.

Stoddart, T. (1992). Los Angeles Unified School District intern program: Recruiting and preparing teachers for an urban context. *Peabody Journal of Education*, 67(3): 84-122.

Strauss, R. P., & Sawyer, E. A. (1986). Some new evidence on teacher and student competencies. *Economics of Education Review*, 5(1): 41-48.

Texas Education Agency (1993). Teach for America visiting team report. Austin: Texas State Board of Education Meeting Minutes, Appendix B.