

**EIGHT QUESTIONS ON TEACHER
LICENSURE AND CERTIFICATION:
WHAT DOES THE RESEARCH SAY?**

EDUCATION COMMISSION OF THE STATES

DECEMBER 2005

BASED UPON RESEARCH REVIEWS BY
RMC RESEARCH CORPORATION
AND BY
BEVERLY BUCK AND TRACEY O'BRIEN,
UNIVERSITY OF COLORADO AT DENVER
AND HEALTH SCIENCES CENTER

The contents of this report were developed under a grant to the Education Commission of the States (ECS) from U.S. Department of Education's Fund for the Improvement of Education, award number R215U000010. The contents do not necessarily represent the policy of the Department of Education and should not be viewed as endorsed by the federal government.

© 2005 by the Education Commission of the States (ECS). All rights reserved.

The Education Commission of the States is an interstate compact created by the states, territories and the U.S. Congress that helps state leaders shape education policy.

ECS is pleased to have other organizations or individuals share its materials with their constituents. To request permission to excerpt part of this publication, either in print or electronically, please write or fax the Communications Department, Education Commission of the States, 700 Broadway, Suite 1200, Denver, CO 80203-3460; fax: 303.296.8332; or e-mail ecs@ecs.org.

TABLE OF CONTENTS

Acknowledgments.....	iv
Summary of the Findings.....	v
About This Report.....	xiii
Introduction: Teacher Licensure and Certification, Research and Policy Decisions.....	1
QUESTION 1	6
<i>What kinds of pedagogical knowledge and practice are related to a teacher’s effectiveness in promoting student achievement?</i>	
QUESTION 2	9
<i>To what extent is the selectivity and rigor of teacher preparation programs associated with teaching quality and effectiveness?</i>	
QUESTION 3	11
<i>What is the relationship between verbal ability and a teacher’s effectiveness?</i>	
QUESTION 4	16
<i>Is there empirical evidence for the validity and reliability of tests and methods frequently used in evaluating a teacher’s effectiveness or quality?</i>	
QUESTION 5	22
<i>To what extent is teaching experience associated with teaching quality and effectiveness?</i>	
QUESTION 6	27
<i>To what extent does initial licensure and certification ensure a teacher’s effectiveness?</i>	
QUESTION 7	32
<i>What is the likely impact of raising teacher licensing and certification standards, specifically in raising cutoff scores on state-mandated tests?</i>	
QUESTION 8	36
<i>Is there empirical evidence of differences in the qualifications and performance of teachers prepared through traditional teacher education programs and those prepared through alternative certification programs?</i>	
References.....	41
Glossary	58

ACKNOWLEDGMENTS

This report owes its existence to the support and efforts of many people.

The report is based on a review of the research literature the Education Commission of the States commissioned from RMC Research Corporation. In addition to the RMC Research review, ECS commissioned Beverly Buck and Tracey O'Brien at the University of Colorado at Denver and Health Sciences Center to complete a further review and synopsis of each resource. Dan Goldhaber of the University of Washington, Cassandra Guarino of RAND and Willis Hawley of the University of Maryland critiqued the original research review and offered valuable feedback.

Several ECS staff also provided invaluable assistance. Michael Allen started the research process and initial writing for the current report, as well as completing the first two reports in this series. Additionally, his general writings for ECS were included in the Introduction section. Tricia Coulter wrote a great deal of the text. Suzanne Weiss created the majority of the Summary of Findings and Josie Canales painstakingly proofread the final manuscript. Kymberly Bloom is responsible for the excellent layout and look of the final online and print products.

In addition, Patricia Lauer, of Mid-continent Research for Education and Learning, made a significant contribution to the report as the author of many of the entries in the Glossary.

Finally, and most significantly, this report would not have been possible without the generosity of the U.S. Department of Education and its Fund for the Improvement of Education, which supported the present report, the two other reports in the series and several other related activities. Particular gratitude is extended to Thelma Leenhouts, the project's program officer, for her constant encouragement and interest throughout the project.

SUMMARY OF THE FINDINGS

What follows is a summary of the findings of an extensive review by the Education Commission of the States (ECS) of empirical research on the effectiveness of current approaches to licensing and certifying teachers. The research review focused on eight questions (and several sub-questions) that are of particular interest and concern to policy and education leaders, including:

- The extent to which certain factors – ranging from a teacher’s college grade point average and scores on licensing and aptitude tests, to the selectivity and rigor of his or her preparation program – are associated with teaching quality and effectiveness
- The relative performance of fully certified teachers and those teaching out-of-field or with emergency credentials
- The relative performance of middle school teachers holding K-8 licenses and those holding dedicated middle school or subject-specific licenses
- The potential benefits and drawbacks of raising teacher licensing and certification standards – specifically, raising minimum passing scores on state-mandated tests.

The full report, available at <http://www.ecs.org/TLCreport>, provides a detailed look at what the research says in response to each of the eight key questions and what that response implies for policy, and includes summaries of the 53 studies reviewed.

Eight Questions on Teacher Licensure and Certification: What Does the Research Say? is the last in a report series on teaching quality supported by a grant from the U.S. Department of Education. The first, an in-depth review of research on teacher preparation published in August 2003, is available at <http://www.ecs.org/tpreport>. The second, which was released in September 2005, focused on what the research says about teacher recruitment and retention. It is available at <http://www.ecs.org/trrreport>.

Question 1:

What kinds of pedagogical knowledge and practice are related to a teacher's effectiveness in promoting student achievement?

Only three studies addressing this question met the review criteria. All three focused on the relationship between classroom practices and student achievement on standardized tests, and all relied on teachers to identify the specific classroom techniques used. It should be noted that while self-reporting is frequently used in social science research, results may differ from data gathered through observation.

For several reasons, this research must be considered **inconclusive**. First, the three studies used differing variables. One looked at how the use of small-group instruction and an emphasis on problem-solving skills affected 10th-grade math scores; another, at the impact of small-group discussion and hands-on learning on 8th-grade math and reading scores; and the third, at the relationship between the amount of time a teacher spent on active instruction (presenting or explaining material, providing feedback and whole-class instruction) on reading and math scores at several grade levels.

The findings of the studies also varied. For example, one of the studies found a positive relationship between the use of small-group instruction and student achievement, while another found that technique to be associated with lower achievement scores.

Finally, ECS offers cautions about these results, including that research looking only at a single grade level may not be able to be generalized to other grades (Wenglinsky, 2002), and that the utility of a particular teaching technique may not manifest as an improvement in achievement scores (Goldhaber and Brewer, 1997b).

POLICY IMPLICATIONS

Further research on this question – using consistent definitions and data-gathering techniques, and taking into account the diverse learning styles and/or aptitudes of students at various grade levels – would clearly be useful, although the effort and cost of such research may be prohibitive. Thus, any policies or requirements that directly address pedagogical techniques should be developed and implemented with great caution.

Question 2:

To what extent is the selectivity and rigor of teacher preparation programs associated with teaching quality and effectiveness?

Only two studies addressing this question met the criteria for inclusion in this review. Both studies, using institutional ratings published in *Barron's Profiles of American Colleges*, found the selectivity of a teacher's preparation program to be associated with higher student

achievement. The findings of these two studies – one of which used data from just a single state – constitute only **limited evidence** of a positive relationship between program selectivity and teaching effectiveness.

POLICY IMPLICATIONS

The research findings suggest that closer examination of the key characteristics and features of teacher preparation programs at more-selective institutions would produce information and insights useful in crafting policy governing teacher preparation and professional development.

Question 3:

What is the relationship between verbal ability and a teacher's effectiveness?

Related Questions:

Do other measures of aptitude, such as academic performance or test scores, predict teacher effectiveness? Is certification through the National Board for Professional Teaching Standards (NBPTS) associated with increased teacher quality and effectiveness?

➤ *Verbal ability*

A number of studies over the past three decades investigating the hypothesis that a teacher's verbal ability is positively related to student achievement offer **strong evidence** that it is.

➤ *Other measures of aptitude*

Several studies reviewed for this report found a positive relationship between teachers' academic performance – as measured by college grade point average (GPA), education coursework and SAT or ACT scores – and their effectiveness in the classroom. But these studies used differing dependent and independent variables. In addition, questions have been raised about the validity of GPA as an assessment measure because of the possibility of grade inflation and inconsistent grading scales, which can lead to overestimation of content knowledge.

Thus, the research reviewed for this report should be taken as offering only **moderate support** for the hypothesis that academic performance predicts teacher effectiveness.

➤ *National Board certification*

No studies investigating the association between certification by the National Board for Professional Teaching Standards and teacher quality or effectiveness met the criteria for this review. It should be noted, however, this review looked only at research completed between 1983 and 2003. Some studies completed in the past couple of years support the assertion that

National Board certification is related to increased teacher quality (Goldhaber and Anthony, 2004).

POLICY IMPLICATIONS

While there is some evidence of the predictive value of grade point averages, test scores and other measures of aptitude, licensure and certification systems that rely heavily or exclusively on such measures should not be implemented without further research.

Question 4:

Is there empirical evidence for the validity and reliability of tests and methods frequently used in evaluating a teacher's effectiveness or quality?

Evaluation tests and methods covered in this report include: Praxis tests, National Board for Professional Teaching Standards (NBPTS) certification tests, state licensure exams, principals' ratings of teachers, teacher work sample systems and portfolio systems.

➤ *Praxis tests*

There is **strong support** that Praxis tests, which have been subject to ongoing evaluation by the Educational Testing Service, are valid and reliable.

➤ *NBPTS certification tests*

Two studies – and one review of previously completed studies – met the criteria for inclusion in this report. The findings of this research were **inconclusive**, due to both the small number of empirical studies and the divergence of the findings.

➤ *State licensure examinations*

Published studies of teacher licensure examinations in four states – Colorado, Connecticut, Massachusetts and Pennsylvania – provide **limited evidence** that such exams typically lack relevance, utility and/or reliability.

➤ *Principals' ratings of teachers*

Several studies of this approach to evaluating teachers were eliminated from review because they used the now-defunct National Teachers Examination as the comparison measure to determine validity. One study that did meet the criteria for review found **high correlations** between principal, peer and self-evaluations and students' performance on reading tests.

➤ *Teacher work sample and portfolio systems*

In the one study of teacher work sample systems that met the criteria for inclusion in this report, researchers found that work samples had content validity – reflecting national, state and local standards as well as the research on effective teaching. Evidence of the effectiveness of this approach to evaluating teachers is thus categorized as **limited**.

No studies assessing the validity and reliability of teacher portfolio systems were found.

POLICY IMPLICATIONS

Further research should be undertaken with the goal of gaining clarity on what we want to measure and whether the methods used to do so are reliable and valid. This issue is of particular importance when it comes to high-stakes assessment used for job retention, promotion or compensation.

Question 5:

To what extent is teaching experience associated with teaching quality and effectiveness?

The research reviewed for this report – more than a dozen studies in all – typically focused on the extent to which student achievement, as measured by standardized-test scores, was correlated with the number of years a teacher had been teaching. Several studies used a different approach – for example, taking into account not only the number of years on the job, but also certification levels and other variables, and comparing the classroom performance of novice and expert teachers using ratings by trained outside observers.

Taken together, these studies provide **strong evidence** of the positive relationship between teaching experience and teaching effectiveness.

It is important to keep in mind, however, that some research also suggests that the positive effects of teaching experience in relation to student achievement are not constantly additive, but instead tend to level off after a few years.

Nor should it be overlooked that teachers with the most experience tend not to be the ones teaching students who are at greatest risk of academic failure. This may artificially inflate the apparent association between teaching experience and student achievement.

POLICY IMPLICATIONS

The field would benefit greatly from research that investigates to what extent the superior performance of experienced teachers is attributable. If it is not just a matter of the length

of time they have been on the job and their subsequent adjustment to job stresses and processes, but to the skills and knowledge they have acquired over the years through interaction and collaboration with other teachers, professional development, mentoring and other experiences may help less-experienced teachers have similar results.

Regardless of the current existence of this type of research, however, the justified assumption of the types of experience from which teachers likely benefit suggests the need for greater policy support for programs and practices that provide teachers – throughout their careers – with the time, resources and tools they need to work together and learn from one another.

Question 6:

To what extent does initial licensure and certification ensure a teacher’s effectiveness?

Related Questions:

How does the performance of middle school teachers with a K-8 license compare with those holding a dedicated middle school or subject-specific license? Is there evidence that multi-tier licensure systems improve the quality of teaching?

The research reviewed for this report offers **strong evidence** that students taught by fully certified teachers achieve at higher levels than those with teachers who are certified but teaching out-of-field, or who hold emergency certification. One notable exception was a study by Goldhaber and Brewer (2000) that found students who had teachers with emergency credentials did no worse on achievement tests than those taught by teachers holding standard credentials.

As for the question of the optimal certification for middle school teachers, the body of research is too limited to be considered anything other than **inconclusive**. The single study meeting the criteria for this review (Mandeville and Liu, 1997) showed that middle school students of teachers with secondary certification in mathematics were better able to solve high-level math problems than students of teachers with elementary certification.

Finally, the literature search for this review did not turn up any studies on the impact of multi-tiered licensure systems on teaching quality.

POLICY IMPLICATIONS

Research provides strong support for policies requiring all teachers to be fully certified and teaching in their field. Of course, establishing such a requirement is one thing; being able to actually fill all teaching slots with highly qualified individuals is quite another – particularly in the case of schools and subjects that are difficult to staff. The challenge for

policymakers is to find ways to both recruit and retain quality teachers, and ensure the equitable distribution of those teachers across and within school districts.

As for the two related questions – K-8 versus subject-specific certification for middle school teachers and the impact of multi-tier licensing systems – these are topics of increasing attention and interest, and clearly merit further research.

Question 7:

What is the likely impact of raising teacher licensing and certification standards, specifically in raising cutoff scores on state-mandated tests?

Related Questions:

Would raising the cutoff scores on required teacher tests increase teacher quality? Would raising these cutoff scores change the demographic makeup of the teaching force?

Some people see raising minimum passing scores on licensing and certification tests as a relatively easy way for states to increase teacher quality. The research reviewed for this report – only two studies, which defined and assessed “teacher quality” differently – provides **limited support** for this hypothesis.

At the same time, several other studies included in this review provide **moderate support** for the claim that raising cutoff scores would lead to a decrease in the diversity of the teaching force.

POLICY IMPLICATIONS

The nature and extent of the relationship between teacher quality and scores on tests used for certification and licensure clearly warrant closer and more rigorous study.

It also is important for policymakers to recognize that any beneficial effects of raising cutoff scores – improved teacher quality, however it is defined and assessed – might be outweighed by the side effect of reduced diversity in the teaching force.

Question 8:

Is there empirical evidence of differences in the qualifications and performance of teachers prepared through traditional teacher education programs and those prepared through alternative certification programs?

Because of the ever-increasing interest in alternative certification programs as a means to draw more teachers into the field, there are burgeoning numbers of programs classified as “alternative

certification.” The amount of variation in requirements and structure among these programs argues against referring to them categorically.

The studies meeting the criteria for inclusion in this report provide **moderate evidence** that teachers prepared through alternative certification programs do not differ from those prepared through traditional teacher education programs in terms of academic qualifications. But the results are **inconclusive** as to differences in their performance in the classroom or the achievement test scores of their students.

POLICY IMPLICATIONS

Alternative certification programs provide an important option for individuals who want to become teachers, and a means for bringing larger numbers of people into the profession. Such programs often are targeted toward attracting potential teachers from underrepresented ethnic or racial groups, or to increase teacher supply in high-demand fields and underserved geographic areas.

Research comparing the quality and effectiveness of traditionally and alternatively certified teachers is limited, at best, and offers little guidance to policy governing how teachers are prepared. The field would benefit greatly from studies incorporating finer-grained variables – such as the timing and structure of student-teaching experiences – and including student achievement as an outcome measure.

ABOUT THIS REPORT

This is the final report in a series of three reports about the research on teaching quality that the Education Commission of the States (ECS) produced through a grant from the U.S. Department of Education's Fund for the Improvement of Education. The focus of this report is on teacher licensure and certification. The first report in the series, *Eight Questions on Teacher Preparation: What Does the Research Say?*, was completed in July 2003. It can be viewed online at <http://www.ecs.org/tpreport> and a print version purchased from ECS at that same Web address. The second report, *Eight Questions on Teacher Recruitment and Retention: What Does the Research Say?*, was completed in September 2005. It can be viewed online at <http://www.ecs.org/trrreport>.

The reports are intended to guide policymakers, educators and foundation officials in their efforts to improve the quality and supply of America's teacher workforce. ECS also hopes the reports will help researchers and others strengthen the knowledge base that underlies policy and practice, and ensure research in the field better addresses the needs and interests of practitioners and, especially, policymakers.

Among ECS' constituents – governors, legislators, state school chiefs and other political and education leaders – the issue of teaching quality consistently ranks as one of their top concerns. This is no doubt due in part to the shortage of well-qualified teachers faced by virtually every state to one degree or another. It also is due to the persuasive and growing body of evidence that teacher effectiveness is the single most-important educational factor in children's achievement in school. Without reliable guidance and the ultimate success of efforts to strengthen teacher quality and supply, however, policymakers and education leaders may turn their attention away from this issue, in spite of its fundamental importance, and pursue other strategies for improving education.

It is hoped this report and the other two in this series can indeed begin to offer the information so greatly needed. This report presents an assessment of the current baseline of the research knowledge relating to specific questions about teacher licensure and certification. As research continues, the report will need to be revised and updated periodically to reflect new studies that may shed light on the questions under consideration here or on other questions about teacher licensure and certification that may emerge over time.

The report also indicates where there is insufficient research to answer the questions asked. This not only has implications for efforts to ground policy decisions in solid evidence but also for the assessment of what additional research needs to be undertaken to provide stronger evidence and more satisfactory answers.

How To Read the Report

The report is structured around eight questions, each of which can be read independently of the others. The discussion of each question includes a narrative review of research addressing that question and offers some policy recommendations based on this review.

In addition to the specific research reviewed for each of the eight questions, the report includes other material to enhance the understanding of the reader. The Introduction provides an overview of the issues involved in teacher licensure and certification, and discusses the role of research in policy decisions. This section also includes a section discussing general considerations about the research reviewed for this report. A general discussion about improving the research on education, including suggestions for roles various stakeholders can play in such an effort, is found in the first report of the series, *Eight Questions on Teacher Preparation: What Does the Research Say?* at <http://www.ecs.org/tpreport>.

Because the report deals with highly technical issues and material, the use of technical terms was unavoidable. Terms relating to research are italicized in colored text (i.e., *term*). Except in the summaries of individual research studies, however, they are noted only the first time they appear in a given section of the report. Holding the cursor over any one of the identified terms causes a pop-up box to appear with a basic definition of the term. Double clicking on the identified term causes a window to appear with the more complete Glossary definition. The Glossary also can be viewed independently.

This report may be used in conjunction with *A Policymaker's Primer on Education Research: How To Understand, Evaluate and Use It*, which ECS and Mid-continent Research for Education and Learning (McREL) developed jointly, to help policymakers and others understand the subtleties of scientific research and be more confident in assessing and using it. The Primer, which was written by Patricia A. Lauer, is accessible online at <http://www.ecs.org/researchprimer> and available in an abridged version at that Web address.

This report notes those instances, via the use of a colored asterisk (*) followed by red text, where the Primer can provide the reader with a more in-depth understanding of the related methodological issues.

The Basis for the Report

The review of the research literature on teacher licensure and certification presented in this report was commissioned by ECS from the RMC Research Corporation. RMC Research employed rigorous criteria in the selection and analysis of the studies they reviewed; the criteria are summarized in the next section. The review presented here represents a summary of what was identified as the more rigorous and reliable research published during the 20 years prior to the completion of the review – research published between 1984 and 2003. The original review was based on 16 questions that were further distilled into the eight questions that compose this report. The 16 questions were chosen through interviews of policymakers and education leaders

conducted by Michael Allen, formerly the program director of the Education Commission of the States Teaching Quality Policy Center.

In addition to the RMC Research review, ECS commissioned Beverly Buck and Tracey O'Brien at the University of Colorado at Denver and Health Sciences Center to complete a further review and synopsis of each resource. Both reviews inform the resulting review of research contained in this report.

How Was the Research Selected?

Researchers at RMC Research Corporation selected the research included for review in this report, and the report relied on their judgment as to the appropriate inclusion criteria. All the literature reviewed for the present report are examples of *empirical research* – studies that offer evidence for their conclusions based on observation rather than articles based on opinion or that use other studies for support. Non-empirical pieces can be quite helpful in clarifying issues conceptually, but since this report addresses empirical questions, it seeks to provide empirical evidence.

RMC Research ultimately selected 105 studies for inclusion in their review, out of 258 articles and book chapters considered for inclusion. After determining the specific areas on which to focus for the current report, that number was further reduced to the 53 studies included in this review. Of the original 258 resources reviewed, a number of potential candidates were eliminated either because they were non-empirical or lacked the characteristics of sound scholarship. The criteria that were used for selection of studies included:

- *Direct relevance* to the questions to be investigated (the questions directly related to the topic at hand and the measures were properly defined)
- Publication in a journal or scholarly book that used *independent peer review*
- Publication by a research organization with a sound reputation for conducting high-quality research and with well-established peer-review processes (only including those that were nonpartisan and who used *quantitative* designs that satisfied the other criteria)
- *Empirical results* that offered quantitative evidence (rather than offering opinions, theories, principles or frameworks)
- *Rigorous methodologies* that met generally accepted standards in relevant research traditions.

Meta-analyses and reviews of the research also were included if they met the criteria and if they added new information. Summaries of the literature were generally not included.

Standards for rigor were:

- *Adequacy of design*: The design must have been developed to answer specific questions, describe how participants were selected for inclusion, operationalize terms, and present enough information to show the design was appropriately and objectively implemented
- *Representativeness of data*: Studies included were specific about sampling frames and the populations to which the results could generalize, and reported the response rate and issues that may have arisen from a low rate.

- *Sound data analysis*: Studies must have used acceptable analytic techniques, controlling for the influence of variables that may *bias* results and acknowledging any limitations to the techniques employed.

Reasonable and unbiased interpretation of results: Studies should have discussed alternative interpretations of the results that were found and/or raised any issues around the *reliability* and *validity* of results associated with these studies.

While the present review cannot claim to be exhaustive, it is hoped it includes virtually all the highest-quality relevant literature published from 1984 through 2003. A complete list of the sources reviewed for this report appears in the References section.

How Was the Research Evidence Assessed?

Assessing how well the research responds to the eight key questions is tricky. The reader will note frequent observations throughout this report about the implications or limitations of the research. These observations often draw on the assessments provided by the researchers at RMC Research in their original research review.

This report attempts to provide an overall evaluation of how strongly the body of studies relevant to a specific question points to a particular answer. How to undertake such an overall evaluation of the research is a subject of intense scientific discussion in and of itself. Even among research methodologists who consider only *quantitative research*, there are disagreements about proper procedure. When, as in the present case, there are both quantitative and *qualitative research* involved, and when there is little *experimental research* that stands above the rest in identifying cause-and-effect relationships, an assessment of the strength of the research base is that much more difficult.

Because the primary purpose of this report is to provide an assessment of the relevant research for policymakers, the designations of the strength of the research are intended to be utilitarian. The criteria employed in making these judgments are certainly not the only ones possible. Hopefully, however, the criteria used here provide a reasonable comparative evaluation and a practical and comprehensible shorthand indication for policymakers who want to use the research evidence in making policy decisions.

The designations of the strength of the research support used in answering the eight questions are as follows:

- The research was considered to offer **strong support or evidence** for a conclusion if (1) there were several solid *experimental studies* or *quasi-experimental studies* that supported it; and/or (2) there were a significant number of *correlational studies* that supported it involving advanced statistical approaches such as *regression analysis*; and (3) there were very few, if any, studies that cast doubt upon the conclusion. In other words, there needed to be an unequivocal pattern of support for the conclusion on the basis of solid quantitative research.

- The research was considered to offer **moderate support or evidence** for a conclusion if it did not meet the criteria for strong support, but (1) there were one or more solid experimental studies or quasi-experimental studies that supported it; and/or (2) there were more than several correlational studies that supported it involving advanced statistical approaches; (3) there were few studies that cast doubt upon the response; and (4) in borderline cases, especially if there was disagreement among studies, there were *simple descriptive studies* present that made it more plausible that certain correlations were based upon a true causal relationship. In other words, there needed to be a clear pattern of support for the conclusion on the basis of solid quantitative research.
- The research was considered to offer **limited support or evidence** for a conclusion if it did not meet the criteria for moderate support, but (1) there was at least one solid experimental study or quasi-experimental study that supported it; and/or (2) there were several correlational studies that supported it involving advanced statistical approaches; (3) there were a preponderance of simple descriptive studies that supported it, and (4) there was considerably weaker evidence in support of any conflicting conclusion.
- If the research for any conclusion did not at least meet the standard of providing limited support, then it was regarded as being **inconclusive**. This could be the case both when only one or two studies supported a conclusion and when there were not significantly more studies that support one conclusion than support one or more opposing conclusions.

Notes About the Research Reviewed

Selection Biases

It is important to note that relying only on published literature invites a bias in favor of research that is of interest to an academic or philanthropic audience and that supports traditionally held positions. Also, it excludes a good deal of the local research and evaluation studies that teacher educators or other researchers conduct in relative obscurity. In general, however, the value of *peer review* is it screens out work of inferior quality and work that has a strong advocacy, rather than scientific, orientation. Moreover, a good deal of local research relies on a set of experiences and assumptions that are often not widely shared outside a local context, so the wider significance or *external validity* of such local studies is often very limited. Finally, it would require an enormous amount of time (and a significantly greater expense) even to locate literature that is either not published or published in a venue other than a peer-reviewed journal. Thus, the restriction of the review to published peer-reviewed literature gives it at least an initial assurance of quality and seemed a reasonable and cost-efficient limitation.

Reviewing only published research carries additional cautions, as well. First, there is a likely “publication bias” that is related to the notion that studies with findings of “no effect” were less likely to be published or offered for publication. Second, given the large number of studies contained in various literature databases and the imperfect functioning of key words as search tools, it may be studies that contain information pertaining to the topics of this review, but not explicitly described as such, were overlooked. Finally, in applying selection criteria, it was relatively easy to distinguish studies that should be included from those that should not, but the studies that met the criteria for acceptance varied in quality. While this report includes notes as

to flaws, readers are still cautioned to bear in mind that variations in quality exist among the studies included for review.

Methodological Concerns

Multiple methodological concerns have been identified in several of the reviews of research in this field and by the RMC Research Corporation in completing the review for this report. Following is a brief discussion of the more salient issues.

Use of *Proxy* Variables

The research generally relies on proxy variables to measure teaching expertise. For example, the assumption was made in many of the studies that if a teacher had a certificate or an advanced degree, he/she had mastered important and/or pre-specified content knowledge and skills. This is not necessarily the case, however. What is taught may not be equivalent to what is learned and/or transferred into practice. Further, the content of what is taught in one teacher education program may vary dramatically from another, thus the knowledge and skills gained from participation in one institution does not necessarily equate to knowledge and skills learned in another. More direct measures of expertise are needed.

Validity and *Reliability* of Survey Data

Many of the studies conducted within the teacher effectiveness literature relied on single point-in-time surveys. These surveys represented a snapshot from a given day. It was unclear the extent to which survey responses would remain stable over time. Many of the survey items were very general and imprecisely measure instructional processes.

Estimation of *Effect Sizes* and Need for Control of Extraneous *Variables* That Influence Learning

The “effect-size” research also reveals substantial difference in reported findings because of the ways in which outcomes were conceived and measured. For example, researchers that used achievement status (i.e., one point-in-time assessment of achievement as measured by a standardized test) were, in fact, examining effects of cumulative experiences on a student, not just the effects of a single year of experience in a classroom. Many of these analyses did not control for the effects of student background, prior achievement and other variables known to influence student achievement and generally, therefore, tended to overestimate effects. Results differed when researchers investigated changes in student achievement over a single year or over multiple years. While these analyses can adjust for effects of background variables and prior student achievement and therefore can produce an effect size that was more likely to reveal true classroom effects, the analyses still contained errors given the known unreliability of gain scores. Thus, according to some researchers (Rowan, Correnti and Miller, 2002), these gains analyses tended to underestimate effects.

Causal Inferences

An additional challenge within the literature is how to interpret the results and the degree to which causal inferences can be drawn. Some patterns from the literature that have been shown to be reliable, for example, still do not account for potential confounding variables such as exposure to advanced curricula or different pedagogies. Few experimental studies have been conducted. Experimental designs, however, are not the “magic bullet” because they, too, face limitations when applied in educational settings. For example, in education it is difficult to control field conditions, especially over long periods of time. Students are mobile, leading both to attrition and contamination of data. Techniques are being developed to address these challenges, but as yet, no method yields consistently reliable data. Rowan and colleagues (2002) explained that an alternative to some of the earlier effect-size approaches is to first estimate “true” rates of academic growth and then assess teacher influence on growth. Early results from researchers using this “explicit growth” method suggest that effect sizes are not similar to those found in the other models. Other research has identified the utility of more recent analytic techniques, such as *hierarchical linear modeling* (HLM), that help control the nested nature and influence of students and teachers in classrooms, schools, districts and states to tease out the influence of system variables on student performance.

Interaction Effects

Only a few of the studies were able to examine interaction or combined effects of multiple influences on teacher effectiveness. This was partially due to the use of large pre-existing data sets that did not measure these variables and by the limitations on the data collected. Combined effects may change the results that were achieved since many either serve directly or synergistically to cancel the apparent effect of other variables.

Location in History

While this research was restricted to studies conducted within the past 20 years, it is possible that the conditions of education have changed sufficiently enough in the current standards-based environment as to prohibit generalizability of early studies. Schools of education have changed, pressures on teachers have increased, and the character of the teaching workforce has changed. These factors may limit the extent to which findings of studies can be combined or generalized to current settings.

Definitional Issues

Many researchers in the field do not use common definitions or measurements for the same variables. For example, K-12 student achievement has been variously measured by scores on state assessments, the National Assessment of Educational Progress, or ACT scores. Teachers’ content expertise has been defined by varying cutoff scores, passage rates of different tests, proxies as explained previously and portfolio assessments. Teacher effectiveness has been measured as implementation of what was learned; self-assessments of comfort, confidence and competence; principals’ assessments of performance; and student test scores. Once again, *syntheses* of results across these measures must be interpreted with caution.

Units of Analysis

Several researchers identified issues in the research from reviewers and those conducting studies mixing up different units of analysis. Some believed that individual students and classrooms were the only appropriate unit of analysis. Others wrote there was value in analyzing data from different levels as long as the analyses were methodologically appropriate for the level.

Need for Caution

Given the limitations of the research, the summary presented here must be viewed with caution. For the literature to become more reliable and of greater utility, it will be important to conduct studies with larger numbers of students and to interpret the results according to a meaningful theory that explains what is found. Some researchers have questioned whether this type of research should be continued, since analyses only reveal what was effective and less effective, but not why. Others have suggested the standards-based environment from the past several years made generalizability of previously completed studies difficult, if not impossible. Nonetheless, the results presented here provide information that can be used to suggest what has worked in the past and policies that may work in the future.

- * **For additional insight into the methodological issues involved in the preceding discussion, see the section titled “How Do I Know if the Research Is Trustworthy?” in *A Policymaker’s Primer on Education Research* found online at <http://www.ecs.org/researchprimer>.**

INTRODUCTION: TEACHER LICENSURE AND CERTIFICATION, RESEARCH AND POLICY DECISIONS

The Critical Importance of Teacher Licensure and Certification

It is generally recognized that – apart from home and other environmental influences – teaching quality has the greatest impact on student achievement. It is important then for states to ensure all students have quality teachers. One method by which a threshold for this type of quality is established is through teacher licensure and certification.

The No Child Left Behind Act of 2001 reinforces the necessity of teacher quality as defined through licensure or certification by requiring states receiving funds under Title I to have a “highly qualified” teacher in every classroom by the 2005-06 school year. A “highly qualified” teacher, as defined by the legislation, must be fully licensed or certified by the state and must not have had any certification or licensure requirements waived on an emergency, temporary or provisional basis. Teachers also must demonstrate subject-matter competency.

All states currently require teachers to have a bachelor’s degree for full certification and to have completed a teacher preparation program. Some states also require teacher candidates to take tests designed to assess their mastery of pedagogical skills and/or subject matter.

There is, however, debate over the value of states’ teacher certification programs and procedures. While proponents claim that fully certified or licensed teachers are often more capable educators, opponents argue that certification does not guarantee competency and serves as an unnecessary obstacle for otherwise well-qualified individuals who wish to enter the teaching profession.

Certification and licensure requirements vary considerably from state to state. Once certified, teachers in most states must renew their certification or license periodically to ensure they are knowledgeable about new developments in their field. In the past, teachers needed only to accrue a certain number of continuing education credits or perhaps earn a master’s degree (in any subject) to maintain licensure. Increasingly, however, states are taking measures to ensure continuing certification requirements motivate teachers to pursue more directed, research-proven career-growth activities.

In addition, states are beginning to require some sort of demonstrated performance as a requirement for continuing licensure or certification. Many states also are aligning requirements for continuing certification with standards for high-quality professional development and standards for exemplary teaching. Some states will grant recertification credit for a master’s degree only if it directly enhances the teacher’s content knowledge or teaching skill.

Another growing trend is “staged licensure,” which confers a limited-time beginning or provisional license to new teachers who pass the requirements for initial certification, a regular or “professional” license to teachers when they demonstrate successful teaching performance,

and then may grant an advanced or “master” license to teachers who demonstrate high levels of accomplishment. In some states, a teacher who receives certification from the National Board for Professional Teaching Standards automatically qualifies for the highest level of licensure.

The teacher licensing and certifying authority itself also varies from state to state. Whereas initial licenses in some states are granted by the college and university programs that prepare teachers, in other states the department of education grants all licenses, while still other states have established a separate and autonomous credentialing or licensing agency.

This level of variability in teacher licensure and certification structures and processes highlight the importance of ensuring policies governing this arena are informed by quality research. The type of research that would be most helpful would cover various aspects of licensure and certification and issues involved in the processes. Primary among these issues is whether and which factors have utility for teacher effectiveness. A number of factors are often appealed to as assurances of teacher effectiveness. One such factor involves academic characteristics, such as a teacher’s verbal aptitude, success in college courses or the selectivity of the undergraduate institution he or she attended. As discussed in this report, the importance of these characteristics varies. It must be noted, however, that while these characteristics are considered important or possibly indicative of future effectiveness, their relevance for policy is more difficult to see. Teacher licensure and certification policies do not usually take institution selectivity into account. As it is presumed to be important, however, it also is important that research be conducted to investigate the accuracy of such suppositions.

Other questions also have arisen about the current system of licensure and certification. For example, if licensure and certification is intended to ensure a threshold of effectiveness, how well does the current system accomplish this? What would be the potential results for the teaching force if the requirements for licensure and certification were raised? The issue of licensure and certification requirements also should be investigated empirically. It would be important to know what the empirical support is for the importance of experience, especially considering that many types or levels of certification are reliant on experience. Policymakers and education leaders also should consider whether there is evidence for the validity or reliability of the assessment tests and methods used to evaluate teachers as some of these figure into licensure and certification systems and decisions.

Finally, the issue of alternative versus traditional certification is receiving increasing attention. Many policymakers and education leaders want to know whether there is a difference in effectiveness between teachers certified through traditional routes and alternative routes. This question, possibly more than others, highlights the importance of understanding all aspects of a question as it is addressed in research. The term “alternative routes” does not describe a homogeneous group of programs. Alternative routes to certification vary widely in their structure, requirements and processes. Therefore, research that categorizes programs as alternative or traditional without providing detailed descriptions of what constitutes the programs themselves may not offer information of any true utility or significance.

The research investigating these and other questions related to teacher licensure and certification is reviewed in this present report. ECS has no vested interest in any particular position on the

issues related to teacher licensure and certification. As much as possible, this report attempts to provide a neutral and objective assessment of the research findings. If there are any acknowledged biases in this effort, they are (1) a desire to find importance for policymakers and others in the body of research reviewed and (2) a concern, on the other hand, not to pretend that the research supports more than it legitimately does.

The Role of Research in Policy Decisions

Policy decisions in education are never made solely on the basis of objective information. There are always values that come into play and, in the world of politics, compromises to win support or bow to fiscal constraints. In addition, education research is never adequate to justify the adoption or development of a particular policy, strategy or program.

There are several reasons for this inadequacy. First, policy decisions often require a commitment of money and resources. The fact that the research provides evidence for the effectiveness of a particular kind of program or strategy does not mean that program or strategy is affordable or cost effective or that it can be supported politically.

One can imagine, for example, licensure and certification systems requiring extensive probationary periods and intensive individualized assessment of pedagogical technique and skill by a team of researchers prior to determining certification status and placement of teachers. This would be an extremely costly, resource intensive and impractical method to put into effect. Additionally, such requirements may have the undesired effect of discouraging qualified and quality individuals from pursuing teaching as a career.

In a similar vein, although research may show that the result of implementing a particular kind of strategy is *statistically significant*, it may not be *practically significant*. The system mentioned above, for example, would be associated with such a large resource and economic price tag as to make it virtually impossible to implement. Also the overall gain in student achievement – the *effect size* – may be slight compared to the effect of engaging in smaller and more manageable changes to a certification system thereby rendering the implementation of such a system unnecessary when weighed against the outcome.

Second, policies, programs and interventions in education are highly contextual, and their success generally depends on the convergence of a number of factors that may not be easily *replicated* or that may not be identified in the research as important to the outcomes observed. In addition to research evidence, then, policymakers or educators need to have good information or else take a leap of faith that the adoption of a policy or program proven successful in one setting also will be successful in another.

Despite these limitations, research contributes valuable information for policy decisions. The weight of research evidence, and especially a lone research study, is never a sufficient guide for policy decisions, but decisions that fly in the face of a sizable body of good research are likely to be ineffective and possibly even disastrous. Also, while not even a whole body of research on a

particular question will provide definitive answers, the verdict of multiple research studies should be regarded as the most reliable guide available.

- * **For additional insight into the methodological issues involved in the preceding discussion, see the section titled “How Do I Know if the Research Warrants Policy Changes?” in *A Policymaker’s Primer on Education Research* found online at <http://www.ecs.org/researchprimer>.**

Considering the Whole Body of Evidence

Decisions about practice or policy should be informed by the entire body of good research available. Proponents of one point of view or another may be able to point to a single study or a number of studies that support their position, while ignoring those that do not. Such a selective use of research cannot provide real assurance the course of action the proponents recommend is wise. Even if the preponderance of research supports a particular decision or policy, evidence to the contrary should not be ignored.

The importance of evaluating the entire body of relevant evidence, as opposed to relying on a single study, holds for fields like health care or agriculture as much as for education. In health care, for example, new findings about the benefits or dangers of certain pharmaceuticals or foods or about the effectiveness of various diets appear with confusing frequency. If a person were to base decisions about what drugs to take or what foods to eat on the findings of each new study, that individual would be changing medications and diet constantly – so frequently, in fact, there would be insufficient time for the true impact of any particular change to be measured. Thus decisions about one’s diet or pharmaceutical prescriptions must be based on an assessment of all available evidence, and apparent conflicts between the findings of different research studies must be explained to the satisfaction of the physician and patient.

The same holds true for education. While new studies about a particular strategy may not appear with the frequency of new research in health care, the investment in any strategy – especially if it is meant to be enacted in policy – is sufficiently great that any change of course will be costly and repeated changes unaffordable. Thus, it is in the best interest of policymakers, educators and other stakeholders to look at the entire body of available evidence when making policy decisions. The more good research that exists, the more it becomes possible to understand the limitations of any individual study and the inconsistencies that may seem to exist between the findings of one piece of research and another. Additionally, in education research, the effect of any change will likely take a substantial period of time to manifest. Changes in policy or practice that are made with an expectation for quick change will likely result in disappointment, frustration and the cessation of efforts that, over time, may have proven worthwhile.

To be sure, it is entirely conceivable – in education as in other fields – a new research study will provide dramatic and powerful new evidence for or against the efficacy of a particular strategy. Until the findings of that study can be confirmed independently by other studies, however, and until the entire body of relevant studies can be reassessed in light of these new findings, the costs, risks, dislocations and other inconveniences that accompany change may make it prudent

to stay the course. On the other hand, in cases where a current practice is demonstrably inadequate or downright harmful, the risks of implementing a new strategy, even though unproven, may be outweighed by the urgent need to make a change.

Question 1:

What kinds of pedagogical knowledge and practice are related to a teacher's effectiveness in promoting student achievement?

What the Research Says

Only three studies were found that met the criteria for inclusion in this report and addressed this issue. Additional studies were pulled for review but not included in this report usually because the studies used *proxies* for teacher ability – often test scores – rather than gathering data on actual pedagogical techniques used. Other reports were not included because they used teacher motivation or expectations as the *independent variable*. While these variables may affect student performance, they do not relate to the present question.

The three studies reviewed below all related classroom practices to student achievement on standardized tests. The use of certain classroom techniques was by teacher self-report. It is important to understand that, while self-report is frequently used in social science research, the results may differ from those same data gathered through observation.

Using the categorical definitions for this report, the results of this research must be considered **inconclusive**. This is due to several factors. First is the small number of reports. This challenge is exacerbated by the fact that each study used different variables. Goldhaber and Brewer (1997b) looked at the reported level of control teachers had over teaching techniques, the use of small group instruction and an emphasis on problem-solving skills and how these pedagogical factors related to 10th-grade mathematics scores. Wenglinsky (2002) also looked at the use of small-group discussion and included hands-on learning, but used a different grade level – 8th grade – and scores on both mathematics and science assessments. Finally, Rowan, Correnti and Miller (2002) used time a teacher spent in active instruction (e.g., whole-class instruction, presenting or explaining material and providing feedback) and its effect on multigrade reading and mathematics scores.

The findings of the studies also varied. Wenglinsky (2002) found a positive relationship between the use of small-group instruction and student achievement. Goldhaber and Brewer (1997b), however, found that technique associated with lower achievement scores.

The final factor involved in finding this research inconclusive is this report's cautions about their results. These include the recognition that utility of a teaching technique may not manifest as an improvement in achievement scores (Goldhaber and Brewer, 1997b) and research looking only at a single grade level may not be able to be generalized to other grades (Wenglinsky, 2002).

SUMMARY OF STUDIES

Three studies that met the criteria for inclusion in this report addressed this issue:

1. **Goldhaber and Brewer, 1997b**, used *production function analysis* techniques to analyze the National Education Longitudinal Study (NELS) data from 1988 to estimate the impact of observable and unobservable schooling characteristics on student outcomes. Relevant to pedagogical skill or practice, teacher practices, including the level of control teachers have over their teaching techniques, teaching in smaller groups and emphasizing problem-solving techniques had an effect on 10th-grade mathematics scores. The researchers found that students with teachers who had little or no control over their teaching techniques demonstrated lower achievement on test scores. This same result was found for teachers who used the other practices mentioned above (small groups and problem solving). [Note: The researchers caution this outcome may not indicate that these teaching techniques are not useful, but simply that they may not improve achievement on standardized tests.]
2. **Rowan, Correnti and Miller, 2002**, used *hierarchical linear modeling* to analyze data from *Prospects: The Congressionally Mandated Study of Educational Opportunity*, a large scale study of schools that served economically disadvantaged children and youth, to see if patterns of active teaching were related to classroom-level differences in students' academic growth. The measures were taken from three sets of questions on a teacher survey. One question asked teachers about average minutes per week spent on instruction in reading and mathematics. The second asked teachers about time spent in active teaching formats (e.g., presenting or explaining material, leading discussion and providing feedback). The third asked teachers the percentage of time students spent in individualized and whole-class instruction. The researchers hypothesized that student achievement would be related to the amount of time the teacher spent as an agent of active instruction – in active teaching formats and with students in whole-class instruction. Generally, this hypothesis was supported: reading and mathematics achievement were positively related to active instruction. An exception was that time spent in individualized instruction had no significant effect on mathematics achievement.
3. **Wenglinsky, 2002**, used multilevel *structural equation modeling* to analyze data from the 1996 National Assessment of Educational Progress (NAEP) for 7,146 eighth graders who took the mathematics assessment and 7,776 eighth graders who took the science assessment to determine the relative and interactive effects of teacher inputs, teacher professional development, classroom practices, class size and student characteristics on student achievement. Wenglinsky found that teachers' use of specific classroom practices (small-group instruction, hands-on learning) had statistically significant relationships with student achievement scores. In particular, when teachers made use of hands-on activities to illustrate concepts in mathematics and science, students performed better on assessments in these subjects (70% of a grade level). When teachers focused on conveying higher-order thinking skills, particularly those that involved strategies to solve different types of problems, students performed better on mathematics assessments. Professional development activities in hands-on learning and higher-order thinking skills also were associated with improved student performance. [Note: Wenglinsky cautioned that the study only covered students at one grade level and two subjects; that the study was *cross sectional* and not *longitudinal*; and that better *proxies* used to measure constructs may be available.]

What It Means for Policy

Because the results of the research reviewed for this question are inconclusive it is inappropriate to draw any clear implications for policy. Additional research using consistent and clearly defined definitions and data-gathering techniques could further the field. As mentioned, however, in the discussion of The Role of Research in Policy Decisions in the Introduction section, the effort and cost to complete this type of research may be prohibitive.

Additionally, research completed in this field needs to consider potential development differences between students at different grade levels that may affect the efficacy of particular techniques. Likewise, issues of classroom diversity and the potential differences in learning styles consequent to this diversity also should be taken into account.

The best conclusion is any policy or requirement that directly addresses pedagogical techniques should be developed and implemented with great caution due to challenges in properly assessing their use and efficacy given the diversity of today's classrooms.

Question 2:

To what extent is the selectivity and rigor of teacher preparation programs associated with teaching quality and effectiveness?

What the Research Says

Only two studies were found that met the criteria for inclusion in this review and addressed this question. Both studies used the ratings of institutions published in *Barron's Profiles of American Colleges* as the indicator of selectivity. Both studies found higher ratings of an institution's selectivity were associated with higher student achievement. Regardless of this consistency in findings, however, the research is taken **inconclusive**. This is due to the small number of studies meeting criteria for inclusion in this report and the fact that one of the studies (Lankford, Loeb and Wyckoff, 2002) used data from only a single state.

SUMMARY OF STUDIES

Two studies met the criteria for inclusion in this report and addressed this issue:

1. **Ehrenberg and Brewer, 1994**, used *multiple regression analysis* and *econometric methods* to examine the extent to which the teacher preparation school and teacher characteristics influenced the probability of public school student achievement and student dropout rates. Data were from the 1980-82 High School and Beyond Longitudinal Survey. The study used a *sample size* of 8,400 students who completed surveys and math, vocabulary and reading tests during their sophomore and senior years. The study used a smaller sample of 2,650 of these students whose teachers completed a 1984 survey on teacher intelligence, verbal aptitude and the name of the institutions at which they received their bachelor's degree. The institutions from which the teachers graduated were rated on a six-point scale using *Barron's Profiles of American Colleges* ratings of the selectivity of admissions requirements. The selectivity of undergraduate institutions attended by high school teachers was positively correlated with students' base-year gain scores, especially for African-American students.
2. **Lankford, Loeb and Wyckoff, 2002**, was a simple *correlational study* that analyzed data from several different sources on every teacher in New York State (approximately 180,000 annually) between 1984-85 and 1999-2000 to determine the variation in the average attributes of teachers across New York public schools. The study examined teacher characteristics associated with student performance. Core data came from the Personal Master File of the Basic Education Data System of the New York State Education Department. Teachers were classified according to whether they had prior teaching experience, a bachelor's degree, certification, teaching "in field," and passage/failure on their first attempt at the National Teacher Examination general knowledge exam or on the New York State Liberal Arts and Science Exam. The study also examined the rating of the undergraduate institution from which the teacher obtained his/her degree using the *Barron's Profiles of American Colleges* rating, classified by "most," "less" and "least competitive" schools. The study found that teachers with Bachelor's degrees from the least competitive

colleges were significantly more likely to have 4th-grade students who were identified as performing at below basic levels on the New York State English Language Arts Exam. Conversely, teachers from the most competitive colleges were significantly more likely to have no students performing below basic levels on that exam.

What It Means for Policy

The question of the effect of an institution's selectivity, and whether and how that should be weighted for its graduates, continues to be present in discussions of teacher quality. While the findings of the studies reviewed imply some support that an institution's selectivity is related to later student achievement, it is difficult to conceive of an appropriate policy recommendation based on these findings. It would be more informative to determine the characteristics of the teacher preparation programs at selective institutions and compare those to characteristics at institutions rated as less selective. This information could then be used to inform policy governing teacher preparation program development. Additionally, it could inform the types of subjects, methods or pedagogical courses and experiences that in-service teachers may find beneficial if offered through professional development.

Question 3:

What is the relationship between verbal ability and a teacher's effectiveness?

RELATED QUESTIONS:

Do other measures of aptitude, such as academic performance or test scores predict teacher effectiveness? Is certification through the National Board for Professional Teaching Standards (NBPTS) associated with increased teacher quality and effectiveness?

What the Research Says

➤ *Related to verbal ability*

Since the Coleman report in 1966, researchers and practitioners have been interested in investigating the *hypothesis* that teachers' verbal abilities are positively related to student achievement. Verbal abilities in these studies were usually measured through a teacher's score on a verbal aptitude test. The majority of this research was completed prior to the previous two decades, the time period covered by the present report, therefore the reviews addressing this issue included below are reviews of previously completed research. With that said, there is **strong support** that a teacher's verbal ability is related to student performance.

➤ *Related to academic performance*

Research completed on other measures of teacher aptitude as it relates to teacher effectiveness presents some challenges. First, there are a variety of methods by which aptitude or ability is assessed. The most frequent measures used in the research reviewed below are college grade point average (GPA), education coursework completed and scores on the SAT or ACT. Cautions are often included when using GPA as an assessment measure because of the possibility of grade inflation and inconsistent grading scales, which can lead to overestimation of content knowledge based on those measures thereby rendering conclusions about their effect misleading.

The second challenge in coming to a conclusion about the research reviewed is the variation in how teacher effectiveness is measured. In two of the studies reviewed below (Ferguson and Womack, 1993; Guyton and Farokhi, 1987) teacher performance is used as the *dependent variable*. When considering any subjective measure, however, it is important to be aware of how the data were gathered. For the studies reviewed below, performance was assessed via raters watching teachers teach. Both studies found a *positive relationship* between education coursework and GPA and teacher performance.

Finally, three studies (Gitomer, Latham and Ziomek, 1999; Latham, Gitomer and Ziomek, 1999; Olsen, 1985) reviewed below looked at scores on various academic measures (e.g., GPA, scores on SAT, ACT or other college admission exams) and education majors. Olsen (1985) found that education majors had higher scores on academic measures. Gitomer et al. (1999) and Latham et al. (1999) broke education majors into two groups – those seeking academic-subject licensure

and those seeking non-academic subject-specific licensure such as elementary licensure. The researchers found higher college admission tests were associated with individuals seeking academic-subject licensure.

While the studies reviewed are strong, the variation in independent and dependent variables used, and questions about the validity of some of the measures (i.e., GPA) lead to a conclusion that the research reviewed below offers only **moderate support** as to whether these measures of academic performance predict teacher effectiveness.

➤ *Related to National Board certification*

No studies investigating the association between certification by the National Board for Professional Teaching Standards (NBPTS) and teacher quality or effectiveness that met the criteria for inclusion in this report. It is important to note, however, the research review completed for this report looked only for research completed between 1983 and 2003. Recently some studies have been completed supporting the assertion that National Board certification is related to increased teacher quality (see Goldhaber and Anthony, 2004).

SUMMARY OF STUDIES

➤ *Related to verbal ability*

Three meta-analyses and research reviews met the criteria for inclusion in this report and addressed this issue:

1. **Greenwald, Hedges and Laine, 1996**, conducted a *meta-analysis* of 29 of Hanushek's studies and 31 other studies published in journals and books that used *production function* approaches. The meta-analysis examined the effects of per-pupil expenditure, teacher quality, class size and teacher salaries on student achievement. Greenwald, Hedges and Laine concluded that greater resource inputs were associated with higher achievement: each type of input had a positive effect. They further concluded that verbal ability rather than the degrees teachers earned or their experiences had the strongest effect on student learning.
2. **Hanushek, 1989**, in a review of 187 separate studies in 38 published articles and books related to expenditure relationships in schools found that the closest thing to a consistent conclusion across the studies was the finding that teachers who perform well on verbal ability tests do better in the classroom, usually measured through scores on tests.
3. **Verstegen and King, 1998**, conducted a review of research studies done to investigate the relationship between resource inputs into schooling and student outcomes as measured by achievement tests. The researchers found that one of the most frequently analyzed teacher characteristics was verbal ability and in 12 of 15 early (1970's) studies, teachers' verbal ability consistently predicted student achievement.

➤ *Related to academic performance*

Five studies met the criteria for inclusion in this report and addressed this issue:

1. **Ferguson and Womack, 1993**, used *multiple regression analysis* to assess the degree to which education and subject matter coursework, grade point average (GPA) and National Teacher Examination (NTE) specialty scores predicted teacher effectiveness. Teaching performance was assessed using a survey instrument completed by supervisors and other faculty after classroom observation. Instructional competence was measured according to 13 categories of expertise based on an index of indicators used to evaluate all teachers. Evaluations of 266 Arkansas Tech University student teachers who were teaching secondary school over the course of seven semesters were collected. Additional evaluations were conducted following employment as classroom teachers for one year. The researchers found that coursework in teacher education made a positive difference in teaching performance and that education coursework was a more powerful predictor of teaching effectiveness than GPA in the major and NTE specialty scores.
- 2,3. **Gitomer, Latham and Ziomek, 1999**, and **Latham, Gitomer and Ziomek, 1999**, conducted a *multiple regression analysis* to determine the degree to which teacher test scores are related to the academic and demographic profiles of a pool of prospective teachers. The study examined the relationship between undergraduate grade point averages (GPAs) and demographic data, SAT and ACT college admission test scores, and scores of over 300,000 prospective teachers who took the Praxis I or Praxis II College of Education entrance examinations and teacher licensure tests. Researchers also examined candidates' undergraduate grade point averages and demographic background. The researchers found teacher academic abilities varied by type of licensure sought. Those individuals pursuing licensure in academic-subject areas had the highest college admissions test scores, while this in non-academic fields (e.g., elementary education) had the lowest.
4. **Guyton and Farokhi, 1987**, conducted a simple *correlational study* to determine whether basic skills and successful academic performance in teacher education programs were associated with subject-matter knowledge and teaching performance. The sample included over 600 graduates of Georgia State University between 1981 and 1984 with scores on the Regents' Test (basic skills) and either the Teacher Certification Test (TCT, basic knowledge) or the Teacher Performance Assessment Instrument (TPAI, teaching competencies). The Regents Test was used to assess reading and writing competencies. Academic quality was defined as performance on two statewide tests, the TCT, and grade point averages. The researchers found higher GPAs were associated with higher scores on the TCT. Scores on the TCT did not predict scores on the TPAI. Additionally, scores on the TCT were related to competencies that deal with the planning stages of teaching, but not related to whether the teacher demonstrated an understanding of the subject matter being taught and could demonstrate its relevance. Guyton and Farokhi concluded there was no significant relationship between performance on a subject-matter test and teaching behavior, but it did find a relatively strong relationship between GPA in education courses and teaching performance. This may suggest that demonstration of knowledge using a paper and pencil test is different from the ability to demonstrate knowledge in an actual teaching situation.

5. **Olsen, 1985**, was a simple *comparative descriptive study*. The researcher collected data on 107 education graduates from the University of Wisconsin-Parkside who completed degrees during the time period from 1981 to 1983 and compared their high school percentile rank, English placement score, math placement scores, cumulative grade point averages (GPA) at graduation, grades for introductory university courses and certification levels with those of 1,420 non-education graduates from the same university who did not have teaching certificates. Education graduates were found to have higher cumulative university GPAs, higher high school percentile ranks and higher grades in the introductory English 101 course. While statistically non-significant, scores on measures related to mathematics favored the non-education graduates. [Note: Olsen pointed out the GPA information should be interpreted with caution since grade inflation in the school of education was possible.]

One literature review addressed teacher testing and met the criteria for inclusion in this report:

1. **Mitchell, Robinson, Plake and Knowles, 2001**, summarized the research on teacher testing for certification for the National Research Council of the National Academy of Sciences. The committee considered whether current teacher licensure tests measure teacher competence appropriately and in a technically sound manner; whether teacher licensure tests should be used to hold states and institutions of higher education accountable for the quality of teacher preparation and licensure; and how innovative measures of beginning teacher competence could help improve teacher quality. The committee reviewed a sample of widely used teacher licensure tests developed by ETS (Education Testing Service) and found they meet most of the criteria for technical quality, although there was room for improvement. Evidence from multiple studies was reviewed, and the group concluded that even well-designed tests could not measure all of the prerequisites of competent beginning teachers. At the time of the study, teacher education had no agreed-upon definitions of competent beginning teachers. Most of the licensing tests that were reviewed were of sufficient technical quality. No conclusions could be drawn on the technical qualities of the National Teacher Examinations because too little data were available. Little research had been conducted to understand the extent to which current teacher licensure tests related to teacher effectiveness. Current tests were found to rely almost exclusively on content knowledge. Comparison of passing rates across states was considered misleading because there was such a wide variability of program characteristics that such comparisons were invalid.

➤ *Related to National Board certification*

No studies were found that met the criteria for inclusion in this report and addressed this issue.

What It Means for Policy

As mentioned previously in the report, one function of licensure and certification is to serve as an indication of teaching quality or effectiveness. Nevertheless, because teachers become certified or licensed before they have substantial teaching experience behind them, indicators of effectiveness incorporated into licensure and certification procedures need to be predictive.

Based on the research reviewed there is moderate support that measures such as grade point average (GPA), the successful completion of education coursework and scores on aptitude tests are related to student achievement. There is likely utility, therefore, in including these measures in licensure and certification procedures. More research, however, should be completed before implementation of a system that relies heavily or exclusively on these measures. The research reviewed in the current report offers only moderate support for the predictive values of these measures, and there are myriad other factors, including pedagogical skill and knowledge, that impact a teacher's effectiveness.

An additional danger in relying on a cutoff score for aptitude tests as a measure of quality is where to draw the line. It would be difficult to argue, for example, that a difference of 50 points on an 800-point test indicates a true difference in teaching ability, particularly a difference in teaching ability that could not be alleviated through experience or inservice training.

Question 4:

Is there empirical evidence for the validity and reliability of tests and methods frequently used in evaluating a teacher's effectiveness or quality?

What the Research Says

The utility of any evaluation tool or procedure is necessarily dependent on the extent to which it is reliable and valid. For this reason, it is important that such evaluation methods have strong empirical support showing they demonstrate these characteristics.

Evaluation tests and methods covered in this report include: Praxis tests, National Board for Professional Teaching Standards (NBPTS) certification tests, state licensure exams, principal ratings of teachers, teacher work sample systems and portfolio systems.

> Praxis tests

Praxis tests have been subject to ongoing *validity* and *reliability* tests by the ETS (Educational Testing Service). Porter, Youngs and Odden (2001) reviewed these studies. There is **strong support** that the Praxis tests are valid and reliable.

> NBPTS certification tests

Two studies and one review of previously completed studies were found that met the criteria for inclusion in this review and investigated some aspect of reliability or validity of the NBPTS certification tests. The findings of this research were **inconclusive**, both due to the small number of *empirical studies* and the divergence of the findings.

One study (Bond, Smith, Baker and Hattie, 2000) that looked specifically at *construct validity* found that National Board certified teachers scored higher on measures of teacher excellence than did noncertified teachers. Another study (Burroughs, Schwartz and Hendricks-Lee, 2000), however, found that NBPTS candidates had difficulty with tasks associated with the certification standards. Finally, a research review completed by Porter, Youngs and Odden (2001) estimated that measurement error led to almost 20% of candidates not achieving certification who had the same qualifications as those who did.

> State licensure examinations

Data on the validity and reliability of state licensure examinations for Colorado, Massachusetts, Pennsylvania and Connecticut were published in the research literature. These studies found many state licensure examinations lacked relevance, utility or reliability. These findings led to the conclusion there is **limited support** the state licensure exams reviewed may not be reliable or valid. The categorization of the support as limited is because so few studies were found for each exam that met the criteria for this report.

➤ *Principal ratings of teachers*

A few studies were found that investigated the relationship of principals' ratings of teachers. Only one study is included in the current review, however, because the other studies used the now-defunct National Teachers Examination (NTE) as the comparison measure to determine validity. Therefore, research using this instrument cannot be considered of any utility. The study (Gallagher, 2002) that met the criteria for inclusion in this report and addressed this issue found high correlations between teacher evaluations by principals, peers and self and reading gain scores of children. As only one study was found that met the criteria for inclusion and addressed this issue the research is considered **inconclusive**.

➤ *Teacher work sample systems*

While many descriptions of teacher work sample systems exist, only one study was found that met criteria for inclusion in this report and addressed this issue. In this study (Denner, Salzman and Bangert, 2001) researchers found teacher work samples had *content validity*, representing national, state and local standards and the research on effective teaching. The findings are categorized as offering **limited support**, however, because only one study was found.

➤ *Portfolio systems*

While Wilkerson and Lang (2003) pointed out that portfolios were in place in nearly 90% of schools, colleges, and departments of education, no studies were found that assessed the validity and reliability of portfolio systems.

SUMMARY OF STUDIES

➤ *Praxis tests*

One research review met the criteria for inclusion in this report and addressed this issue:

1. **Porter, Youngs and Odden, 2001**, in their review of *psychometric* studies of teacher assessments, reported that Praxis tests have been subject to ongoing validity and reliability tests by the ETS (Educational Testing Service). One study showed that 93% of assessors had no need to reconcile ratings with other assessors since the *interrater reliability* was so high. In another survey, assessors rated the Praxis criteria 3.5 on a 5-point scale of comprehensiveness.

➤ *NBPTS certification tests*

Two studies and one research review met the criteria for inclusion in this report and looked at this issue:

1. **Burroughs, Schwartz and Hendricks-Lee, 2000**, conducted a *qualitative study* of four candidates for the National Board of Professional Teaching Standards (NBPTS) certification

to determine the ways in which portfolio tasks were perceived and the ease or difficulties they faced as they translated their knowledge and engaged in a “discourse community.” *Stratified purposive sampling* was used to select teachers who provided contrasts of certification areas and teaching sites. Study data included group field notes, observations, interviews focusing on teacher backgrounds, attitudes toward NBPTS and experiences in the portfolio process. The researchers found that all candidates had difficulty representing their experiences in writing. They were apprehensive about writing, had trouble representing tacit knowledge and understanding sampling logic, and struggled with negotiating the NBPTS standards and providing evidence in their teaching. Burroughs et al. suggested that alternate discourse styles, whether generated by culture or context, might affect both NBPTS scorers and candidates.

2. **Bond, Smith, Baker and Hattie, 2000**, investigated the *construct validity* of the National Board for Professional Teaching Standards certification tests. The researchers performed an intensive comparative examination of data from 65 teachers from two certificate areas: Early Adolescence/English Language Arts and Middle Childhood/Generalist. Evidence analyzed included teachers’ instructional objectives and lesson plans for a given instructional unit, observational visits to all 65 teachers’ classrooms, and scripted interviews of the teachers and their students. Of the 34 teachers with Early Adolescence/English Language Arts certification, 13 were National Board certified (NBCT) and 21 were not (Non-NBCT). Of the 31 teachers with Middle Childhood/Generalist certification, 18 were certified and 13 were not. The groups were compared along 15 dimensions, including student work product. In every comparison between NBCTs and Non-NBCTs on the dimensions of teaching excellence, NBCTs obtained higher mean scores. In 11 of the 13 comparisons, these differences were highly *statistically significant*.
3. **Porter, Youngs and Odden, 2001**, reviewed studies that have been conducted to establish overall *reliability* of each of the National Board of Professional Teaching Standards assessments. The estimates indicated that because of measurement error about 19% of individuals who took the test were not certified at the same level of qualifications as those who passed. About 10% of those who were certified probably had fewer skills than those who did not pass.

➤ *State licensure examinations*

Four studies met the criteria for inclusion in this report and looked at state licensure examinations:

1. **Cobb, Shaw, Millard and Bomotti, 1999**, conducted analyses to determine the existence of various types of *validity* for the Program for Licensing Assessments for Colorado Educators (PLACE) test developed by National Evaluation Systems in 1998. Analyses occurred over seven iterations of the test from October 1994 to October 1996. Using *regression analysis* and *analysis of variance*, the researchers analyzed 11,390 test scores representing 5,588 preservice teacher candidates from five Colorado teacher preparation institutions. The researchers found the general content design of PLACE was well aligned with other testing reform taxonomies. The researchers, however, noted that *construct validity* was problematic

at least for those from culturally diverse backgrounds. Cobb et al. also concluded the Basic Skills test battery lacked relevance and utility in many cases, and that some of the content area tests needed to be rewritten to increase *content validity*.

2. **Haney, Fowler, Wheelock, Bebell and Malec, 1999**, used *multiple regression analysis* and *qualitative research* to determine the accuracy of the Massachusetts Teacher Test (MTT) in assessing the reading and writing skills of the test-takers, using data from state and academic reports from 1998. There were insufficient data available to test *concurrent validity*, so an independent review panel examined *reliability* on the April and July administrations of the test using test-retest procedures. Data from 219 teachers who took tests during both periods were analyzed and correlations were established for those people who took the subtests at each time period. The researchers found the scores on the reading and writing MTTs were highly unreliable, with a margin of error close to double, and sometimes triple, the range found on well-developed tests. Thus, a person taking the test multiple times could have huge score fluctuations even if his or her skill level did not change significantly. The study also found the MTT contained questionable content, which made it a poor tool for measuring the test-taker's reading and writing skills.
3. **Popham, 1992**, examined 32 licensure examinations during the period from 1975 to 1991 that had been developed by various organizations, universities and states to determine their *content validity*. Popham found licensure tests varied in their focus on opportunity to learn relative to job relevance and in the stringency of the directions to scoring panelists. For example, some panelists were told to judge by relevant content while others were told to judge by "appropriately measured necessary content." Variations also were found in the ways in which panelists' ratings were analyzed or content quality was computed. Some used majority vote while others averaged the scores. Indices of high, medium and low varied widely.
4. **Wylie and Tannenbaum, 2003**, was a *simple descriptive study* to examine the job relevance of 70 knowledge and skill indicators in five content areas (literature, fine arts, mathematics, social studies and science) for the development of the Pennsylvania beginning teacher licensure program. A random sample of 1,700 teachers and 300 teacher educators were selected to participate in the study, and 626 usable surveys were returned. Respondents judged each of the indicators on the test on a 5-point scale of importance. Literature, mathematics, social studies and science received mean ratings of at least "moderately important;" the mean rating for fine arts fell below "moderately important." Only one-third of the individual knowledge and/or skill indicators within the domains were judged as important (job relevant) enough for consideration in the development of the licensure assessment.

➤ *Principal assessment*

One study met the criteria for inclusion in this report and addressed principal assessment:

1. **Gallagher, 2002**, used *hierarchical linear modeling*, to investigate the relationship of teacher evaluation to student achievement. The sample consisted of 34 elementary school teachers in a charter school. Test scores for 584 students from grades 2-5 who had at least two years of Stanford Achievement Test scored (spring 2000 and spring 2001) were examined. Teachers were evaluated by principals and a peer, and also provided a self-assessment. An average score was derived from the three scores for each domain area. Domains included lesson planning, classroom management, literacy, mathematics and language development. Results indicated a moderately high relationship between teacher evaluation and student achievement in reading. No relationships were found between teacher evaluations in mathematics or language arts and student achievement. *Qualitative* results suggested the different relationships between teacher evaluation and student achievement in the subject areas may be due to more pedagogical knowledge of teachers and evaluators in reading than in math, and better alignment between standards and assessments in reading compared with math. [Note: ECS cautions the *sample size* was low, the data were restricted to a single year of growth for students.]

➤ *Teacher work sample systems*

One study met the criteria for inclusion in this report and dealt with teacher work sample systems:

1. **Denner, Salzman and Bangert, 2001**, conducted a study to examine the *validity* and generalizability of the Teacher Work Samples to assess teachers' abilities to meet national and state teaching standards, and to improve student achievement. A range of candidates was recruited to participate in the study, including junior-level candidates, teaching interns, experienced teachers and teachers who had National Board certification. Researchers collected 132 work samples comprising a description and analysis of teaching and learning context, achievement targets, assessment plan, instructional sequence of at least six learning activities over four weeks, analysis of student learning, and evaluation and reflection. There were strong associations between ratings of teacher work and analysis of student learning based on student work. Ratings also clearly distinguished between teachers on a developmental continuum, although there were no associations with the experience level of the teacher that submitted the sample. The samples had *content validity*, representing national, state and local standards and the research on effective teaching.

What It Means for Policy

As with so much of the other research areas reviewed for this report, the limited or inconclusive nature of the findings necessitate caution in policy recommendations. Methods and systems of evaluation – either through testing and self-report or observational review – are what is relied on in most fields to assure competence and determine quality. If the measure or system is not valid or reliable, any assessments made based on it also are rendered invalid and useless. For this reason, it is important these tests and systems measure what they are supposed to measure and do so consistently. Further research should be undertaken with the goal of gaining clarity on what is being measured and whether the methods by which these variables are measured are doing so.

This issue is of particular importance when it comes to high-stakes assessment used for job retention, promotion or compensation. The best policy recommendation is likely that no assessment measure or method should be required until and unless it has been shown to be reliable and valid.

Question 5:

To what extent is teaching experience associated with teaching quality and effectiveness?

What the Research Says

The research investigating the effects of teaching experience on teacher effectiveness tended to correlate the number of years a teacher had been teaching with student achievement as measured through standardized test scores. Some researchers used years of experience and certification as a *proxy* for expertise and compared novices to experts using ratings by trained outside observers. Different researchers investigated the phenomenon in different ways. Some correlated student test scores with teacher experience while others divided groups using some cutoff number of years (such as five years) and examined differences between these groups.

The results for the research reviewed below offer **strong support** for the benefit of teaching experience for student achievement, specifically after the first few years on the job. Some caveats, however, need to be mentioned to accurately understand the research on teacher experience as it relates to student achievement. First, the effects of teacher experience tend to level off after the first few years (Rivkin, Hanushek and Kain, 2002; Ferguson, 1991), so it is inappropriate to assume a constant additive effect of years of teaching experience. Additionally, teachers with the most experience tend not to teach students at greatest risk of academic failure. This may artificially inflate the apparent association between teaching experience and student achievement.

SUMMARY OF STUDIES

Eleven studies and five *meta-analyses* or research reviews met the criteria for inclusion in this report and addressed this issue:

1. **Ehrenberg and Brewer, 1994**, used *multiple regression analysis* to examine the extent to which school and teacher characteristics influenced student dropout rates and student achievement of those who did not dropout. Data were from the 1980-82 High School and Beyond Longitudinal Survey. The sample was 2,650 students who completed surveys and math, vocabulary and reading tests during their sophomore and senior years and whose teachers completed a 1984 survey on various teacher characteristics. Greater teacher experience was associated with higher base-year scores on achievement tests for white students, but no association between teacher experience and test scores was found for African American or Hispanic students. Also, a higher percentage of teachers in a school with 10 or more years of experience was associated with lower dropout rates for African American students.
2. **Ferguson, 1991**, used *multiple regression analysis* to analyze student data covering 900 districts and 2.4 million students in Texas for the school years 1985-86, 1987-88 and 1989-90 and teacher scores on the Texas Examination of Current Administrators and Teachers (TECAT), a state recertification exam required of all Texas teachers in 1986. The researcher

found teacher experience accounted for over 40% of the variance among student test scores. Teachers with more years of experience produced higher student test scores, lower dropout rates and higher rates of students taking the SAT.

3. **Fetler, 1999**, used *multiple regression analysis* to determine the association between teacher preparation, teacher experience and student achievement. Data were collected from 795 California high schools and over 14% of the 56,571 full-time secondary school mathematics teachers in the state. Data from the 1998 Professional Assignment Information Form were examined to determine demographics, assignments, position and credential status of the teachers. The Stanford Achievement Test, 9th Edition (1998) test scores were used to analyze student achievement. Fetler used Aid to Families with Dependent Children, a *proxy* for poverty, as a *control* in several of the analyses. About half of the teacher sample had 10 or more years of experience and most of the rest had less than five years of experience. Student poverty was found to have the strongest impact on test scores. After controlling for poverty, however, the average number of years of teaching experience was positively related to student achievement in mathematics. [Note: ECS cautions that higher scores for students could potentially be attributed to attrition of lower-performing students or to selective testing.]
- 4,5,6. **Goldhaber and Brewer, 1997a, 1998**, used *multiple regression analysis* and economic *production function analysis* techniques to analyze the National Education Longitudinal Study data from 1988 to determine the extent to which teacher experience and teacher advanced degrees enhanced a teacher's effectiveness in raising student achievement in the areas of mathematics, science, English and history. Tenth-grade student achievement was examined to determine the relative impact of previous test scores, student and family background variables, and variables associated with teacher characteristics and schooling. Researchers found that teacher experience was not related to student achievement. Another analysis of the same data (**Goldhaber and Brewer, 1997b**), however, reported that students with more-experienced teachers had higher scores.
7. **Hawkins, Stancavage and Dorsey, 1998**, was a simple *correlational study* using data from the 1996 National Assessment of Educational Progress (NAEP) Mathematics Assessment. The researchers found that 4th- and 8th-grade students taught by teachers with more than five years of teaching experience outperformed students whose teachers had less than five years of experience.
8. **Okpala, Smith, Jones and Ellis, 2000**, used *multiple regression analysis* to study the impact of selected educational resources (school size, class size, teacher education, teacher experience) and family demographics (participation in free/reduced school lunch and parents with post-high school education) on the achievement scores in reading and mathematics of 4,256 fourth-grade students attending 42 public schools in North Carolina in 1995-96. While all *variables* were significant in some regard, the researchers found the percentage of teachers with 10 years of teaching experience was significantly correlated with mathematics achievement and reading achievement.

9. **Rivkin, Hanushek and Kain, 2002**, used *hierarchical linear modeling* and *production function analysis* techniques to examine data from approximately 3,000 schools and 600,000 students collected as part of the University of Texas – Dallas Schools Project. The study examined the impact of teacher quality and specific other teacher characteristics such as experience and education on student achievement as measured by scores on the Texas standardized state assessments. The researchers found that teacher experience was related to student achievement, with the greater impacts occurring after the first few years of teaching. There was little evidence that the effect of experience continued after that time. Teachers in their first year, and to a somewhat lesser extent, in their second year, tended to perform significantly worse than those with more experience in the classroom. Following the initial period where significant improvement was seen, however, there was little additional improvement over time in terms of impact on measured achievement.
10. **Rowan, Correnti and Miller, 2002**, used *hierarchical linear modeling* to analyze data from *Prospects: The Congressionally Mandated Study of Educational Opportunity*, a large-scale study of student achievement of economically disadvantaged children and youth. The study focused on whether a teacher had special certification to teach reading or math; a bachelor's or master's degree in English (when reading achievement was analyzed) or math; and teacher experience as a *proxy* for teachers' professional knowledge. The researchers found that teacher experience was a small but *statistically significant* predictor of achievement, both for early and later grades. In mathematics, however, there was a positive effect of teachers' experience on mathematics achievement only in the later grades.
11. **Stafford and Barrow, 1994**, was a *comparative descriptive study* that used student test score data collected by the Houston Independent School District from 1985-88 and conducted interviews with principals and administrators to examine differences in student achievement by teachers who had traditional certification compared to those with alternative certification. Other variables, such as teacher experience, also were investigated. Results showed that elementary school students whose teachers had five or more years of experience and interns with teaching experience had *statistically significantly* higher achievement test scores than elementary school students whose teachers with alternative or traditional certification of one year or less experience as a teacher. This relationship was found for all three years of the study. No differences by experience level were found for secondary school teachers.

Five *meta-analyses* or reviews of others' research addressed teaching experience and met the criteria for inclusion in this report:

1. **Greenwald, Hedges and Laine, 1996**, conducted a *meta-analysis* of 29 of Hanushek's studies and 31 other studies published in journals and books that used *production function* approaches. The meta-analysis examined the effects of per-pupil expenditure, teacher quality, class size and teacher salaries on student achievement. Greenwald, Hedges and Laine concluded that greater resource inputs were associated with higher achievement: each type of input had a positive effect. The estimated *effect size* for teacher experience was small. Hanushek, 1996 (see below) disputed the results of this study as unsupported.

2. **Hanushek, 1986**, was a *meta-analysis* of the research on the economics of education and schooling, focusing on production and efficiency aspects of schools. He found that teacher experience has a clear majority of estimated coefficients pointing toward student achievement gains and almost 30% of the estimated coefficients are statistically significant by conventional standards. He further suggested, however, these positive correlations may result from more senior teachers having the ability to select schools and classrooms with better students.
- 3,4. **Hanushek, 1996**, disputed the results of Greenwald, Hedges and Laine (1996, see above), concluding that Greenwald et al. distorted the evidence by using a flawed statistical approach that biased their study by narrowing the number of articles that were included. Hanushek argued that Greenwald et al. defined the inquiry too narrowly, resulting in an overestimate of the effects of resource variables including the teacher variables. The article is based upon **Hanushek, 1997**, a *meta-analysis* of 377 studies, which included 171 studies of the effects of teacher education and 207 studies of the effects of teacher experience. That analysis found that while nearly 30% of the studies measuring teacher experience on student achievement were *statistically significant* in the positive direction, 5% were negative and nearly 66% of the estimated effects were statistically insignificant. In Hanushek, 1996, the researcher stated that, without further analysis, all one could conclude from these statistics is that this particular input is used productively in some circumstances, and policymakers cannot craft effective policy without knowing what distinguishes the significant from the insignificant.
5. **Verstegen and King, 1998**, conducted a review of research studies done to investigate the relationship between resource inputs into schooling and student outcomes as measured by achievement tests. The researchers found that one of the most frequently analyzed teacher characteristics was teaching experience and in 24 of 30 studies years of teaching experience significantly predicted student achievement.

What It Means for Policy

The relationship between teaching experience and student achievement has several implications for policy. The most clear are those dealing with certification or licensure levels. If certification or licensure is intended as an indication of a teacher's effectiveness, an argument could be made as to the importance of including some requirement for experience as part of a system of licensure or certification that includes tiers or levels.

Less obvious potential policy implications can be found by deeper analysis of what is involved in teacher experience. Is experience beneficial simply because teacher's learn intricacies of classroom management and navigation of the system by working within it? If this is the case, then years of employment as a teacher are what matters. Years of experience, however, may actually be a *proxy* for other variables. These variables could include the amount of professional development a teacher has taken over time, the amount of collaboration with colleagues a teacher has participated in through the years, and other such factors related to time on the job. If this is the case, that type of experience may be encouraged and gained more quickly by institutionalizing these types of practices through mentoring and induction practices, creation of

teacher learning communities and the like. These types of practices could be developed and implemented through policy levers. Further and more detailed research into the variables involved in a teacher's experience could shed further light on this issue and its potential implications for appropriate policy and practice.

Finally, any potential policy recommendation or implications must be considered in light of the differences in the distribution of experienced teachers. As mentioned above, experienced teachers are not equally distributed to all schools, specifically to those schools serving students at the greatest risk of academic failure. This inequity in distribution may lead to inaccurate conclusions about the true effects of teacher experience and student achievement.

Question 6:

To what extent does initial licensure and certification ensure a teacher's effectiveness?

RELATED QUESTIONS:

How does the performance of middle school teachers with a K-8 license compare with those holding a dedicated middle school or subject-specific license? Is there evidence that multi-tier licensure systems improve the quality of teaching?

What the Research Says

➤ ***Related to the extent to which initial licensure and certification ensures a teacher's effectiveness***

Researchers investigated this question by looking at certification status or level of teachers and the effects on student achievement on standardized tests. Certification levels were most often categorized as fully certified, certified but without the endorsement for the subject taught (teaching out-of-field) and teaching with emergency certification. The studies reviewed used different units of analysis, addressing data at the national, state, district, school and/or classroom levels.

There was **strong support** that certification level was positively associated with student achievement. A notable exception to the general findings was the Goldhaber and Brewer (2000) study that found students who had teachers with emergency credentials did no worse than students with teachers holding standard credentials.

➤ ***Related to K-8 as opposed to dedicated middle school or subject-specific license***

Scans of databases that typically contain information on articles related to certification and licensure yielded almost no studies that addressed this question. The one study found that met the criteria for inclusion in this report (Mandeville and Liu, 1997) showed middle school students of teachers with secondary certification in mathematics were better able to solve high-level mathematics problems than students of teachers with elementary certification. This may indicate some benefit to subject-specific licensure. Such a conclusion, however, cannot appropriately be drawn from such scant evidence.

➤ ***Related to multi-tiered licensure systems***

As of 2003, at least four states (Arkansas, Connecticut, Kentucky and Wisconsin) had multi-tiered licensure systems (Hill and Dozier, 2003). The literature search for this review did not identify any studies on the impact of multi-tiered licensure systems on teaching quality.

SUMMARY OF STUDIES

➤ *Related to the extent to which initial licensure and certification ensures a teacher's effectiveness*

Seven studies met the criteria for inclusion in this report and addressed this issue:

- 1,2. **Darling-Hammond, 1999**, used *regression analysis* to analyze National Assessment of Educational Progress (NAEP) data for students from 1990 to 1996 and the Schools and Staffing Survey from the same years. The researcher found the proportion of fully certified teachers in a state with a major in the field in which they taught was associated with the strongest achievement scores. These qualifications accounted for 40% to 80% of the variation across states on 4th- and 8th-grade students' scores in reading and mathematics, controlling for students' socioeconomic status and language background. The strongest negative predictor of student scores was the proportion of teachers who were uncertified and the proportion that held less than a minor in the field they taught. The same relationships were consistently found for 1990, 1992, 1994 and 1996 data sets. **Darling-Hammond (2000)** updated this study to include 50 states and found the same results. Whether a teacher had full certification and a major in the subject they taught was positively related to student achievement. These results were robust across all grade levels and subject matters. A state's average NAEP scores were positively associated with the percent of fully certified teachers, and negatively associated with the percent of teachers who were teaching out-of-field.
3. **Fetler, 1999**, used *multiple regression analysis* to determine the association between teacher credential status, teacher experience and student achievement. Data were collected from 795 California high schools and over 14% of the 56,571 full-time secondary school mathematics teachers in the state. Data from the 1998 Professional Assignment Information Form were examined to determine demographics, assignments, position and credential status of the teachers. Variables for determining teacher math skill were full math subject-matter credentials (standardized content test); emergency permits (bachelor's degree, basic skills test and partial coursework in mathematics); limited-assignment emergency permits (valid teaching credential in another subject); or waivers (passing the math portion of a basic skills test). The Stanford Achievement Test, 9th Edition (1998) test scores were used to analyze student achievement. Having a higher percentage of teachers with emergency permits in a school was associated with lower student math scores.
4. **Goe, 2002**, used *multiple regression analysis* to examine the relationship between the percentage of teachers holding emergency permit (EP) teacher certification and California student achievement. The study used the 1999-2000 data from the schoolwide Academic Performance Index (API), which was based on the performance of 6,389 students on the Stanford Achievement Test, 9th Edition, *aggregated* for elementary, middle and high schools. *Variables* included teacher certification type; teacher demographics (EP teachers and first-year teachers); school size; and student demographics such as race/ethnicity, eligibility for free/reduced-price meals and parents' education. There was a direct *negative correlation* between the number of teachers who held emergency credentials and student achievement at the school level. In other words, the greater the number of teachers holding

emergency credentials, the lower the student achievement at the school level. Most of the variation in scores was explained by factors related to student poverty and student ethnicity, but the relationship of emergency credentials to scores existed even when the other factors were controlled.

5. **Goldhaber and Brewer, 2000**, used *multiple regression analysis* and economic *production function analysis* to investigate the relationships between student achievement and type of teacher certification and license. The study used data on about 6,000 twelfth-grade math and science students and more than 2,200 mathematics and science teachers taken from National Educational Longitudinal Study of 1988. Student achievement was measured by examining 12th-grade students' test scores on the Stanford Achievement Test, 9th Edition. Teacher certification types were categorized as probationary or emergency certificates, private school certification, certificates for teaching that were outside the subject area in which the teacher was teaching and advanced degrees. The researchers found teachers with standard certification had a significant positive impact on student test scores compared to teachers with private school certification or who were not certified in their subject area. Unexpectedly, there was no significant difference for mathematics and science test scores for students who had teachers with emergency credentials versus standard credentials.
6. **Hawk, Coble and Swanson, 1985**, used *analysis of variance* to determine the relationship between mathematics teacher certification and student achievement. Student achievement was measured by Stanford Achievement Test scores in general mathematics and algebra of students in grades 6-12 in North Carolina. The researchers compared differences in these scores for students taught by certified teachers with appropriate mathematics endorsements and those taught by certified teachers without this endorsement. It also compared ratings that teachers received on the North Carolina Teacher Performance Assessment System. Results showed *statistically significant* differences in student achievement, favoring teachers who were teaching in their field for both general mathematics and algebra.
7. **Lankford, Loeb and Wycoff, 2002**, was a simple *correlational study* that used data on New York state teachers employed from 1984 to 2000 to determine which schools had the least qualified teachers and whether the distribution of teachers was impacted by attrition and transfer. Core data came from the Personal Master File of the Basic Education Data System of the New York State Education Department. Teachers were classified according to whether they had prior teaching experience, a bachelor's degree, certification "in field," passage/failure on the National Teacher Examination general knowledge exam or the New York State Liberal Arts and Science Exam on their first attempt. The researchers found the proportion of lower-performing students at a school was related to the proportion of teachers at that school who were not certified to teach in any of the subject matters to which they were currently assigned.

➤ *Related to K-8 as opposed to dedicated middle school or subject-specific license*

One study met the criteria for inclusion in this report and addressed this issue:

1. **Mandeville and Liu, 1997**, examined the effects of teacher certification on student mastery of higher-level mathematics concepts. The *sample* consisted of 4,869 students whose teachers had strong preparation in mathematics and 4,492 students whose teachers had little or no formal preparation in mathematics. The sample included students from 266 South Carolina schools with at least one 7th-grade mathematics class. Pairs of schools with similar characteristics, such as student demographics, school location, size and grade organization, were matched. Teachers were classified as high mathematics preparation if they had secondary certification in mathematics and 12 or more credit hours in mathematics beyond initial certification. Teachers who had elementary certification or were teaching out-of-field were considered as having low mathematics preparation. Students responded to 45 items that measured mathematics understanding on the Stanford Achievement Test and 23 items that measured thinking skills. Subskill areas were number concepts, computation and application. *Aggregated* scores for each of the three thinking levels were obtained by deriving number correct per student and averaging across the school. The researchers found middle school students of teachers with more mathematics content-area preparation were better able to solve higher-level mathematics problems than students of teachers with less specialized training. For all three thinking levels, mean mathematics scores for the students of highly prepared teachers were higher than the mean mathematics scores for students of less-prepared teachers. Differences were insignificant for lower level mathematics problems.

➤ *Related to multi-tiered licensure systems*

No studies were found that met the criteria for inclusion in this report and addressed this issue.

What It Means for Policy

The strong support found for the importance of full certification and its positive effect on student achievement warrant the recommendation that all teachers be fully certified and teaching in their field. This recommendation is somewhat moot, however, in light of the requirement that all teachers be highly qualified according to the definitions set out in the No Child Left Behind Act.

Having this type of requirement in legislation, however, and the ability to fill all teaching slots with highly qualified teachers teaching in the subject area for which they hold licensure, certification or endorsement are different issues. This is of particular importance when dealing with schools and subjects that are particularly challenging to staff. This then becomes an issue of recruiting and retaining quality teachers and ensuring equitable distribution of those teachers to all schools and students. Issues of recruitment and retention can be affected by policy (see the previous report in this series, *Eight Questions on Teacher Recruitment and Retention: What Does the Research Say?* at <http://www.ecs.org/trrreport> for a discussion of these issues).

Related to the issue of the type of licensure or certification that may be best for middle school teachers to ensure teacher effectiveness, no conclusion could be drawn on the topic due to the lack of empirical research; therefore, no policy recommendation can be made. The challenge of licensure or certification for middle school teachers is a topic of increasing focus and merits further research.

The complete lack of studies related to the impact of multi-tier licensure systems is interesting. It can be assumed the implementation of such a system is intended to moderate or indicate levels of teaching quality or effectiveness. This is further evidenced by the usual requirement of different levels of experience or education in order to move from tier to tier. While common sense would support this system, common sense is sometimes not supported by research so additional empirical evidence may better inform the necessity of this type of system to meet student achievement or other goals.

Question 7:

What is the likely impact of raising teacher licensing and certification standards, specifically in raising cutoff scores on state-mandated tests?

RELATED QUESTIONS:

Would raising the cutoff scores on required teacher tests increase teacher quality? Would raising these cutoff scores change the demographic makeup of the teaching force?

What the Research Says

The issue of state-mandated testing for teachers to become fully licensed or certified is intended to ensure all certified teachers meet minimum competencies as defined by the state. States using these tests are empowered with setting their own passing cutoff scores. As such it is easy to understand why some people would see raising the cutoff scores for these tests as a relatively easy way to increase teacher quality – by increasing the minimum threshold. The research reviewed below addresses two aspects of this option: whether raising the cutoff scores would increase teacher quality and what impact such an action would have on the teaching force.

➤ Related to raising teacher quality

Two publications (Ferguson and Brown, 2000; Mitchell, Robinson, Plake and Knowles, 2001) addressed the issue of cutoff scores and teacher quality. Both supported the theory that raising the cutoff scores would likely have a positive effect on student achievement or teacher quality. One (Ferguson and Brown, 2000) found that raising scores on the state teacher test was positively related to an increase in student achievement scores. The other (Mitchell et al., 2001) reviewed multiple studies and concluded that raising teacher test scores would likely result in raising the quality of the teaching force. Because of the limited number of publications addressing this issue, the research is taken as providing **limited support** that raising the cutoff scores would result in a consequent increase in teacher quality. In addition to the limited number of qualified studies on the topic, this conclusion also is based on the lack of consistency in defining and assessing quality.

➤ Related to demography of the teaching force

The impact of raising cutoff scores on the demographics of the teaching force was not so positive. Of the four studies that met the criteria for inclusion in this report and addressed this issue, all four predicted a raise in teacher test cutoff scores would result in a decrease in the diversity of the teaching force, sometimes dramatically. The importance of having a diversified field of teachers is almost universally held as an important value and goal, both for the children and society. Because these studies are predicting a possible effect rather than implementing the change and subsequently gathering data, the results are categorized as offering **limited support** that raising cutoff scores would decrease diversity in the teaching force.

SUMMARY OF STUDIES

Five studies and one literature review met the criteria for inclusion in this report and addressed these issues:

1. **Angrist and Guryan, 2003**, used *multiple regression analysis* to estimate the impact of state-mandated certification tests on teacher quality and teacher wages. The researchers used data from the Schools and Staffing Survey from the 1987-88, 1993-94 and 1999-2000 data sets for public schools with over 50 students. Teacher quality was measured by average SAT scores of teachers' undergraduate institutions, whether the institution was a research university or liberal arts college, the proportion of teachers with alternative state certification and proportion of teachers with a degree in the subject they taught. Demographic information of the teachers also was collected. The researchers postulated that any barrier to entry into a field is likely to raise wages. Their finding supported this suggesting that the policy of state-mandated teacher testing was associated with increased teacher wages but not with a corresponding increase in teacher quality. Additionally, Angrist and Guryan found that testing requirements had no effect on the percent of new or inexperienced teachers who were African American or female, but there was a negative association between testing requirements for basic skills and the number of new teachers who are Hispanic reducing the proportion of new teachers who were Hispanic by about two percentage points.
2. **Ferguson and Brown, 2000**, used education *production function analysis* to reanalyze data from Texas and Alabama (see Ferguson, 1991) to show the impact of raising certification scores on student achievement. Using teachers' scores on the 1986 Texas Examination of Current Administrators and Teachers (TECAT) for 900 districts during the 1980s and ACT scores for Alabama, Ferguson and Brown examined whether the differences between students' gain scores from grades 3-5 and grades 9-11 reflected the differences in elementary and high school teachers' scores. The results indicated TECAT scores predicted students' math scores and, additionally, a change in TECAT scores resulted in a change in student-achievement scores over two years.
- 3.4. **Gitomer, Latham and Ziomek, 1999**, and **Latham, Gitomer and Ziomek, 1999**, conducted a *multiple regression analysis* to determine the degree to which teacher testing affects the academic and demographic profiles of a pool of prospective teachers. The study examined the relationship between undergraduate grade point averages and demographic data, SAT and ACT college admission test scores, and scores of over 300,000 prospective teachers who took the Praxis I or Praxis II College of Education entrance examinations and teacher licensure tests. Results showed that teacher academic ability as measured by SAT or ACT scores varied widely by type of licensure sought. Teacher candidates with the highest test scores were more likely to seek licenses in academic-subject areas; teachers with lower scores sought licenses in elementary education and non-academic fields. The higher the Praxis passing score set by states, the higher the SAT and ACT average scores of the passing population. Raising the standards resulted in reducing the pool of candidates in those states and limiting the racial and ethnic diversity of the pool of prospective teachers who met passing requirements.

5. **Murnane, Singer, Willett, Kemple and Olsen, 1991**, used *multiple regression analysis* to investigate the relationship between cut scores and teachers' employment. The researchers examined two sets of data from the National Longitudinal Surveys of Labor Market Experiences database. The first set was the employment history for 20,614 individuals who entered teaching in Michigan public schools from 1972 to 1981 and who were followed through the 1984-85 school year. The second set was for 50,502 teachers who were licensed between 1974 and 1985 and who were followed through the 1985-86 school year. Only individuals with no prior teaching experience who were either white or African American were included in the study. Variables studied included gender, race, academic major, test scores, IQ, subject specialties and year of graduation. Data on district characteristics included proportion of children from low-income families, proportion of professionals, median income, median adult education and changes in licensing requirements over time. Results were based on percentages of the population who received teaching licenses and percentages that entered the teaching profession. Researchers found that with stricter licensing requirements, African American college graduates tended to leave the teaching profession. When North Carolina proposed an increase in the cut score from 644 to 655 on the National Teacher Examination (NTE) Communications Skills Test, the percentage of African American candidates who obtain licenses was predicted to drop from 36% to 5%.

One literature review met the criteria for inclusion in this report and looked at the likely impact of raising teacher certification and licensure standards:

1. **Mitchell, Robinson, Plake and Knowles, 2001**, summarized the research on teacher testing for certification for the National Research Council of the National Academy of Sciences. The committee considered whether current teacher licensure tests measure teacher competence appropriately and in a technically sound manner; whether teacher licensure tests should be used to hold states and institutions of higher education accountable for the quality of teacher preparation and licensure; and how innovative measures of beginning teacher competence could help improve teacher quality. Evidence from multiple studies was reviewed, and the researchers concluded even well-designed tests could not measure all the prerequisites of competent beginning teachers, and while raising test scores would likely result in raising the quality of the teaching workforce, the action would limit its diversity.

What It Means for Policy

Most states require some type of test for teacher licensure or certification. As these tests are intended as a threshold measure for teacher quality, it is understandable that raising the cutoff scores is seen as an option to increase the quality of the teaching force. Before any change in policy in this direction is considered, however, all aspects of the issue should be considered.

First, the research on whether teacher quality would be affected by raising teacher test cutoff scores is limited based on the research reviewed above. This is especially salient when this action could result in a reduction of the diversity of the teaching force, an issue of such great concern that policy and practice have been implemented with the opposite goal.

From the standpoint of basing policy on quality research, this issue is intimately involved with how teacher quality or effectiveness should be measured. Prior to taking a policy action that could have a detrimental effect (i.e., reducing diversity in the teacher workforce) more research should be completed that directly investigates the relationship between teacher quality and/or effectiveness and the relationship of those measures to test scores on specific tests being used for certification and licensure. It is possible that the change in teacher quality, by whatever method it is measured, would not be worth the reduction in diversity thereby making it an untenable step to take.

Additionally, as with all social science research, it is important to recognize there may be confounding variables that better account for a person not passing a teacher test and have nothing to do with that individual's potential quality or efficacy as a teacher. It would be important to be confident in the assessment and data prior to making such a decision.

Question 8:

Is there empirical evidence of differences in the qualifications and performance of teachers prepared through traditional teacher education programs and those prepared through alternative certification programs?

What the Research Says

“Alternative certification” is a general term for nontraditional routes that lead to teacher licensure. These programs are generally geared toward people who already have a baccalaureate degree and would like to become classroom teachers, but require methods coursework and classroom experience to gain certification. Alternative certification programs vary in requirements and can be administered at the federal, state or district levels. Because of the ever-increasing interest in alternative certification programs as a means to draw more teachers into the field, there are increasing numbers of programs classified as “alternative certification.” It is important to note, however, the amount of variation in requirements and structure among these programs makes it difficult, if not impossible, to meaningfully refer to them categorically.

The studies that met the criteria for inclusion in this report and addressed this issue tended to look at academic qualifications and performance in the classroom. Only one study (Laczko-Kerr and Berliner, 2002) used student achievement as an outcome measure. While this study found that students of traditionally prepared certified teachers had higher achievement tests, one study was not adequate to offer conclusions as to empirical support.

There is **moderate support** that teachers prepared through alternative certification programs do not differ from those prepared through traditional teacher education programs in their academic qualifications. The results are **inconclusive** as to differences on measures of performance between teachers prepared through these different routes. The categorization as inconclusive is based on divergent findings and differences in how the *dependent variables* were assessed.

Four of the seven studies reviewed in this report used performance evaluations as a dependent variable. They differed, however, on the findings. Guyton, Fox and Sisk (1991) found no differences in teacher performance. Two other studies (Hawk and Schmidt, 1989; Jelmberg, 1996), however, found that traditionally prepared teachers were rated higher on measures of classroom performance than teachers from alternative certification programs. Finally, one study (Lutz and Hutton, 1989) found differences between the groups depended on by whom they were rated with supervisors rating alternatively prepared teachers higher than traditionally prepared teachers and the opposite true in principal evaluations.

Two studies used differing dependent variables to assess the issue. Houston, Marshall and McDavid (1993) used self-report. While they found that alternatively certified teachers reported greater problems with classroom and work activities (including student motivation, time management and dealing with administration) than traditionally certified teachers, most of these differences disappeared after eight months of teaching.

Finally, Knight, Owens and Waxman (1991) surveyed students for their assessment of the classroom environment. They found that students of traditionally certified teachers reported feeling more challenged, and the classroom was more cooperative and cohesive than students with alternatively certified teachers.

SUMMARY OF STUDIES

Seven studies met the criteria for inclusion in this report and looked at this issue:

1. **Guyton, Fox and Sisk, 1991**, was a *comparative descriptive study* that compared 23 beginning teachers prepared through the 1988-89 Alternative Preparation Institute with 25 beginning teachers prepared through traditional teacher education institutions on teaching attitudes, efficacy, performance and retention in the profession. Surveys measured: whether teaching was student-centered or directive; attitudes toward students, school environment, teaching and support; locus of control; self-confidence; satisfaction with education in society; comfort in school; teaching problems; and teacher efficacy. Open-ended items on the survey probed decisions to become a teacher, influences on adopting teaching as a profession and conceptions about teaching. Two administrators for each teacher completed a 15-item teacher performance evaluation. No differences were found on grade point average, educational attitudes, performance evaluations or teaching attitudes. Teachers with alternative certification were found to be more positive about their improvement in teaching abilities over the month prior to the survey. They also were significantly less satisfied with the structure and organization of education in society. They were less positive than traditionally certified teachers about teaching and staying in the profession, though all but one of the teachers was returning to teach the next year.
2. **Hawk and Schmidt, 1989**, was a *comparative descriptive study* that examined the differences between teachers who entered teaching through traditional programs and those who entered teaching with a degree in a field other than education (“lateral entry”). Scores on the National Teacher Examination (NTE) and the Teacher Performance Appraisal Instrument (TPAI) were compared. Sixteen lateral-entry program candidates, five in mathematics and 11 in science, and 18 traditionally prepared candidates, comprised the sample. There were no *statistically significant* differences on NTE test scores between the two types of teachers. There were differences on the TPAI, favoring the traditionally prepared teachers on four of the five function areas that were measured. The exception was instructional monitoring. *Interrater reliability* on the TPAI ratings, however, was low as was the *sample size*.
3. **Houston, Marshall and McDavid, 1993**, was a *comparative descriptive study* of survey responses from 69 traditionally certified first-year elementary school teachers with 162 alternatively certified elementary teachers in the Houston Independent School District over a two-year period. The study sought to determine whether differences existed in the classroom problems the teachers faced, such as student motivation, burnout and grading procedures, and in their confidence, satisfaction and plans to continue teaching. The researchers found after two months of teaching, alternatively certified teachers perceived significantly greater problems with student motivation, managing time and the amount of paperwork, grading students and dealing with school administration than traditionally certified teachers.

Traditionally certified teachers received greater mentoring assistance than alternatively certified teachers in securing materials and equipment, parent cooperation, student involvement, teaching freedom and peer acceptance. After eight months of teaching, most of the differences disappeared. Traditionally certified teachers were more likely to be female, younger, single and white, and more likely to be teaching in areas where they were certified. Teachers with alternative certification were more likely to be teaching children of color.

4. **Jelmberg, 1996**, was a *comparative descriptive study*. The study used data gathered through a survey of a random sample of New Hampshire elementary and secondary school teachers of mathematics, science, English language arts and social studies who were certified either through traditional or alternative routes to determine differences in teacher performance, motivation to teach, overall preparation and intention to remain in teaching. The final sample consisted of 200 traditionally certified teachers, 30 teachers with alternative certification and 136 principal surveys. The alternatively certified teachers had been certified for two to three years, but had been teaching full time for three years more than that. They had received their certification after completing three years of supervised full-time teaching (with no preservice training) and an approved professional development plan. Results showed no difference in the academic credentials of the two groups, but on almost all other measures, the traditionally prepared teachers fared better than the alternatively certified teachers. These included principals' evaluations of teacher performance and teachers' own ratings of how well prepared they were to teach. Traditionally certified teachers were more likely to have entered teaching because they wanted to work with children, while alternatively certified teachers were more likely to have entered teaching because jobs were available.
5. **Knight, Owens and Waxman, 1991**, was a *quasi-experimental study* using *multiple regression analysis* that examined the relationship between teacher certification type and student perceptions of the classroom environment. The student *sample* consisted of 676 elementary and middle school students and 24 teachers from several public school districts in and around a large city in the American Southwest. Students were divided in two groups: one group had teachers who possessed traditional certification through university- or college-based teachers education programs and who had participated in student teaching; the other group had teachers who had participated in alternate certification programs. None of the teachers had master's degrees or above. The teachers administered an adaptation of the "MyClassInventory" to the students. The inventory consisted of 43 items that measured satisfaction, friction, difficulty, cohesiveness, competitiveness, cooperation, higher- and lower-level thought processes, pacing, homework assignments, and parent involvement. The researchers found students of traditionally certified teachers appeared to be more challenged by their schoolwork but felt the pacing in relation to the demands of the schoolwork was appropriate. They also perceived their classrooms to be more cooperative and cohesive. Students with alternatively certified teachers did not perceive as much opportunity for higher-level thinking and perceived less cohesiveness and cooperation and more friction in their classrooms. [Note: Knight et al. caution that, while these differences were *statistically significant*, there are other teacher characteristics that were not *variables* in this study, which could have contributed to the differences in students' perceptions.]

6. **Laczko-Kerr and Berliner, 2002**, was a *comparative descriptive study* to examine teacher effectiveness based on certification status in five low-income, inner-city school districts in Arizona. The study compared certified teachers (bachelor's degree from an accredited institution and completion of 45 semester hours of elementary education coursework) with "under-certified" teachers, which included emergency-certified (holders of bachelor's degrees from accredited institutions with little or no education coursework) and provisionally certified teachers (some education training but standard certification not fulfilled) and Teach for America teachers. The study measured student-achievement test scores on the Stanford Achievement Test, 9th Edition in the areas of reading, language arts and mathematics for the school years 1998-99 and 1999-2000. The *sample* included 109 matched pairs of under-certified and certified teachers of students in grades 3-8. Art, music and special education teachers were not included in the sample. Results showed that students taught by certified teachers outperformed students taught by "under-certified" teachers by about 20% (or two months on a grade-equivalent scale) in reading, mathematics and language. Teach for America teachers were no more effective than other under-certified teachers.
7. **Lutz and Hutton, 1989**, was a *comparative descriptive study* and *multiple regression analysis* to evaluate the alternative teacher certification program in the Dallas Independent School District, which was designed to prepare teachers for culturally diverse, inner-city schools. The study examined the relationship between teacher characteristics, attitudes and performance, and type of certification. Specific measures included basic-skills test scores, scores on the Texas Teacher Appraisal System, Teacher Advisor Comparison Rating Forms, Teacher Work-Life Inventory, the statewide certification examination, a teacher concerns checklist and a survey of mainstreaming options. The *sample* consisted of 62 traditionally prepared first-year teachers and 110 first-year interns with full-time teaching responsibilities in the district's Alternative Certification Program. Results demonstrated that first-year teachers had higher commitments to teaching as a profession and planned to stay in teaching longer than those who had alternative certifications. Supervisors rated nearly all teachers with alternative certifications (92%) as high as or higher than first-year teachers on performance measures. Principals rated beginning teachers higher than alternatively certified interns on reading, discipline management, classroom management, planning, instructional techniques and instructional models.

What It Means for Policy

There is a great deal of interest in determining whether there are differences between traditionally and alternatively prepared teachers in their quality or effectiveness. Empirical research does not exist, however, that supports whether such differences exist. Additionally, the variability in the structure and requirements among alternative routes to certification make it difficult, if not impossible, to make generalizations about these programs.

Alternative certification programs provide an important option for individuals who want to become teachers, and a method by which a larger number of people can be brought into the profession. Additionally, these programs often are targeted toward attracting potential teachers from underrepresented ethnic or racial groups, underserved geographic areas, or individuals with

subject expertise in high-demand fields (Mikulecky, Shkodriani and Wilner, 2004). These are important values in the teaching field and important goals that are forwarded through these programs.

At the least, the findings of this research review indicate the limits of the evidence about alternative routes to certification. The field would benefit greatly by the completion of research incorporating more fine-grain variables in their comparisons. These variables could include the types of courses taken and the timing and structure of student-teaching experiences. Pulling these variables and their relationship to student achievement and teacher performance would lend more effective guidance to policy governing how teachers are prepared.

REFERENCES

Bibliography of Cited Sources

- Angrist, J., and Guryan, J. (2003, March). *Does teacher testing raise teacher quality? Evidence from state certification requirements*. National Bureau of Economic Research Working Paper 9545.
- Bond, L., Smith, T., Baker, W., and Hattie, J. (2000). *The certification system of the National Board for Professional Teaching Standards: A construct and consequential validity study*. University of North Carolina at Greensboro, Center for Educational Research and Evaluation.
- Burroughs, R., Schwartz, T., and Hendricks-Lee, M. (2000). "Communities of practice and discourse communities: Negotiating boundaries in NBPTS certification." *Teachers College Record*, 102(2), 344-374.
- Cobb, R.B., Shaw, R., Millard, M., and Bomotti, S. (1999). "An examination of Colorado's teacher licensure testing." *Journal of Educational Research*, 92(3), 161-176.
- Darling-Hammond, L. (1999). *State teaching policies and student achievement*. Seattle, WA: University of Washington, Center for the Study of Teaching and Policy.
- Darling-Hammond, L. (2000). "Reforming teacher preparation and licensing: Debating the evidence." *Teachers College Record*, 102(1), 28-56.
- Denner, P.R., Salzman, S.A., and Bangert, A.W. (2001). "Linking teacher assessment to student performance: A benchmarking, generalizability and validity study of the use of teacher work samples." *Journal of Personnel Evaluation in Education*, 15(4), 287-307.
- Ehrenberg, R., and Brewer, D. (1994). "Do school and teacher characteristics matter? Evidence from high school and beyond." *Economics of Education Review*, 13(1), 1-23.
- 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., and Brown, J. (2000). "Certification test scores, teacher quality, and student achievement." In D.W. Grissmer and J.M. Ross (Eds.), *Analytic issues in the assessment of student achievement*. Washington, DC: U.S. Department of Education, National Center for Education Statistics, 133-156.
- Ferguson, R.F., and Womack, S.T. (1993). "The impact of subject matter and education coursework on teaching performance." *Journal of Teacher Education*, 44(1), 55-63.

- Fetler, M. (1999). "High school staff characteristics and mathematics test results." *Education Policy Analysis Archives*, 7(9). Retrieved May 5, 2004, from <http://epaa.asu.edu/epaa/v7n9.html>.
- Gallagher, H.A. (2002, April). *The relationship between measures of teacher quality and student achievement: The case of Vaughn Elementary*. Consortium for Policy Research in Education, CPRE-UW Working Papers Series TC-02-05. Retrieved from May 5, 2004, <http://www.wcer.wisc.edu/CPRE/papers/pdfVaughn%20TE%204-02.pdf>.
- Gitomer, D., Latham, A.S., and Ziomek, R. (1999). *The academic quality of prospective teachers: The impact of admissions and licensure testing*. Princeton, NJ: Teaching and Learning Division, Educational Testing Service.
- Goe, L. (2002). "Legislating equity: The distribution of emergency permit teachers in California." *Education Policy Analysis Archives*, 10(42). Retrieved May 5, 2004, from <http://epaa.asu.edu/epaa/v10n42>.
- Goldhaber, D., and Anthony, E. (2004). *Can teacher quality be effectively assessed?* Seattle, WA: Center for Reinventing Public Education.
- Goldhaber, D.D., and Brewer, D.J. (1997a). "Evaluating the effect of teacher degree level on educational performance." In W. Fowler (Ed.), *Developments in school finance*. Washington, DC: National Center for Education Statistics.
- Goldhaber, D.D., and Brewer, D.J. (1997b). "Why don't schools and teachers seem to matter? Assessing the impact of unobservables on educational productivity." *Journal of Human Resources*, 32(3), 505-520.
- Goldhaber, D.D., and Brewer, D.J. (1998, October). "When should we reward degrees for teachers." *Phi Delta Kappan*, 80(2), 134, 136-138.
- Goldhaber, D.D., and Brewer, D.J. (2000). "Does teacher certification matter? High school teacher certification status and student achievement." *Educational Evaluation and Policy Analysis*, 22(2), 129-145.
- Greenwald, R., Hedges, L.V., and Laine, R.D. (1996). "The effect of school resources on student achievement." *Review of Educational Research*, 66(3), 361-396.
- Guyton, E., and 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.
- Guyton, E., Fox, M.C., and Sisk, K.A. (1991). "Comparison of teaching attitudes, teacher efficacy, and teacher performance of first-year teachers prepared by alternative and traditional teacher education programs." *Action in Teacher Education*, 13(2), 1-9.

- Haney, W., Fowler, C., Wheelock, A., Bebell, D., and Malec, N. (1999, February). "Less truth than error? An independent study of the Massachusetts teacher tests." *Education Policy Analysis Archives*. Retrieved May 5, 2004, from <http://epaa.asu.edu/epaa/v7n4/>.
- Hanushek, E.A. (1986). "The economics of schooling: Production and efficiency in public schools." *Journal of Economic Literature*, XXIV, 1141-1178.
- Hanushek, E.A. (1989). "The impact of differential expenditures on school performance." *Educational Researcher*, 18(4), 45-51.
- Hanushek, E.A. (1996). "A more complete picture of school resource policies." *Review of Educational Research*, 66(3), 397-410.
- Hanushek, E.A. (1997). "Assessing the effects of school resources on student performance: An update." *Educational Evaluation and Policy Analysis*, 19(2), 141-164.
- Hawk, P., Coble, C.R., and Swanson, M. (1985). "Certification: It does matter." *Journal of Teacher Education*, 36(3), 13-15.
- Hawk, P.P., and Schmidt, M.W. (1989). "Teacher preparation: A comparison of traditional and alternative programs." *Journal of Teacher Education*, 40(5), 53-58.
- Hawkins, E.F., Stancavage, F.B., and Dorsey, J.A. (1998). *School policies affecting instruction in mathematics*. Washington, DC: National Center for Education Statistics.
- Hill, T.L., and Dozier, T. (2003, March). *Multi-tiered, performance-based licensure: Four state profiles*. (Policy Brief 5122). Denver, CO: Education Commission of the States.
- Houston, W.R., Marshall, F., and McDavid, T. (1993). "Problems of traditionally prepared and alternatively certified first-year teachers." *Education and Urban Society*, 26(1), 3-34.
- Jelmsberg, J. (1996). "College-based teacher education versus state-sponsored alternative programs." *Journal of Teacher Education*, 47(1), 60-66.
- Knight, S.B., Owens, E.W., and Waxman, H.C. (1991). "Comparing the classroom learning environments of traditionally and alternatively certified teachers." *Action in Teacher Education*, 12(4), 29-34.
- Laczko-Kerr, I., and Berliner, D.C. (2002). "The effectiveness of 'Teach for America' and other under-certified teachers on student academic achievement: A case of harmful public policy." *Education Policy Analysis Archives*, 10(37). Retrieved May 5, 2004, from <http://epaa.asu.edu/epaa/v10n37/>.
- Lankford, H., Loeb, S., and Wycoff, J. (2002). "Teacher sorting and the plight of urban schools: A descriptive analysis." *Educational Evaluation and Policy Analysis*, 24(1), 37-62.

- Latham, A.S., Gitomer, D., and Ziomek, R. (1999). "What the tests tell us about new teachers." *Educational Leadership*, 56(8), 23-26.
- Lutz, F.W., and Hutton, J.B. (1989). "Alternative teacher certification: Its policy implications for classroom and personnel practice." *Educational Evaluation and Policy Analysis*, 11(3), 237-254.
- Mandeville, G.K., and Liu, Q. (1997). "The effect of teacher certification and task level on mathematics achievement." *Teaching and Teacher Education*, 13(4), 397.
- Mitchell, K.J., Robinson, D.Z., Plake, B.S., and Knowles, K.T. (Eds). (2001). *Testing teacher candidates: The role of licensure tests in improving teacher quality*. Washington, DC: National Academy Press. Retrieved November 6, 2003, from <http://www.nap.edu/books/0309074207/html>.
- Murnane, R.J., Singer, J., Willett, J., Kemple, J., and Olsen, R. (1991). (Eds.). *Who will teach? Policies that matter*. Cambridge, MA: Harvard University Press.
- Okpala, C.O., Smith, F., Jones, E., and Ellis, R. (2000). "A clear link between school and teacher characteristics, student demographics and student achievement." *Education*, 120(3), 487-494.
- Olsen, D.G. (1985). "The quality of prospective teachers: Education vs. non-education graduates." *Journal of Teacher Education*, 36(5), 56-59.
- Popham, W.J. (1992). "Appropriate expectations for content judgments regarding teacher licensure tests." *Applied Measurement in Education*, 5(4), 285-301.
- Porter, A.C., Youngs, P., and Odden, A. (2001). "Advances in teacher assessment and their uses." In V. Richardson (Ed.), *Handbook on research on teaching* (4th ed.). Washington, DC: American Educational Research Association, 259-297.
- Rivkin, S.G., Hanushek, E.A., and Kain, J.F. (2002, Revised). *Teachers, schools and academic achievement*. (NBER Working Paper No. 6691).
- Rowan, B., Correnti, R., and Miller, R. J. (2002). *What large-scale, survey research tells us about teacher effects on student achievement: Insights from the Prospects Study of Elementary Schools*. (CPRE Research Report Series RR-051). Philadelphia: University of Pennsylvania, Graduate School of Education. Retrieved May 5, 2004, from <http://www.cpre.org/Publications/tr51.pdf>.
- Stafford, D., and Barrow, G. (1994). "Houston's alternative certification program." *The Educational Forum*, 58, 193-200.
- Verstegen, D.A., and King, R.A. (1998, Fall). "The relationship between school spending and student achievement: A review and analysis of 35 years of production function research."

(ERIC Document Reproduction Service No. EJ575184). *Journal of Education Finance*, 24(2), 243-262.

Wenglinsky, H. (2002). "How schools matter: The link between teacher classroom practices and student academic performance." *Education Policy Analysis Archives*, 10(12). Retrieved May 3, 2004, from <http://epaa.asu.edu/epaa/v10n12/>.

Wilkerson, J.R., and Lang, W.S. (2003). "Portfolios, the Pied Piper of teacher certification assessments: Legal and psychometric issues." *Educational Policy Analysis Archives*, 11(45). Retrieved December 3, 2003, from <http://epaa.asu.edu/epaa/v11n45/v11n45.pdf>.

Wylie, C., and Tannenbaum, R. (2003). *Knowledge and skill indicators important for beginning teachers: A survey of Pennsylvania teachers and teacher educators*. (ETS Research Report RR-03-05). Princeton, NJ: Educational Testing Service.

Other Sources Reviewed

Anderson, M. (1998, November 29). "Value-added puts Tennessee on the map." *The Commercial Appeal*, A18-20.

Angus, D.L., and Mirel, J. (2001). *Professionalism and the public good: A brief history of teacher certification*. Washington, DC: Thomas B. Fordham Foundation. Retrieved May 5, 2004, from <http://www.edexcellence.net/foundation/publication/publication.cfm?id=12>.

Ashton, P., Crocker, L., and Olejnik, S. (1986). *Does teacher education make a difference? A literature review and planning study*. Final report to the Institute on Student Assessment and Evaluation (Grant No. 050-94640-850000). Tallahassee: Florida Department of Education.

Ayers, J.B., and Qualls, G.S. (1979). "Concurrent and predictive validity of the National Teachers Examinations." *Journal of Educational Research*, 73(2), 86-92.

Ballou, D., and Podgursky, M. (1998). "The case against teacher certification." *The Public Interest*, 32(1), 17-25.

Barnes, S., Salmon, J., and Wale, W. (1990, Spring/Summer). "Alternative teacher certification in Texas: A look at initial results." *Teacher Education and Practice*, 6(1), 29-34.

Belden, N., and Plattner, A. (1999). *Teacher quality: A review of existing survey data*. (ERIC Document Reproduction Service No. ED440945). Denver, CO: Education Commission of the States.

Berk, R.A. (1988). "Fifty reasons why student achievement gain does not mean teacher effectiveness." *Journal of Personnel Evaluation in Education*, 1(4), 345-364.

- Berry, B. (2001, May). "No shortcuts to preparing good teachers." *Educational Leadership*, 32-36.
- Bidwell, C.E., and Yasumoto, J.Y. (1999, October). "The collegial focus: Teaching fields, collegial relationships and instructional practice in American high schools." *Sociology of Education*, 72(4), 234-256.
- Bliss, T. (1992). "Alternative certification in Connecticut: Reshaping the profession." *Peabody Journal of Education*, 67(3), 35-54.
- Boe, E.E., and Barkanic, G. (2000). *Critical factors in developing a highly qualified national teaching force*. Paper presented at the American Educational Research Association Annual Conference. New Orleans, LA.
- Bond, L. (2001, May). "On 'Defrocking the National Board': A reply to Podgursky." *Education Next*. Greensboro, NC: University of North Carolina at Greensboro, Center for Educational Research and Evaluation.
- Bond, L., Jaeger, R., Smith, T., and Hattie, J. (2000, Summer). "The certification system of the National Board for Professional Teaching Standards: A construct and consequential validity study." *Education Next*. Greensboro, NC: University of North Carolina at Greensboro, Center for Educational Research and Evaluation.
- Bracey, G., and Molnar, A. (2003). *Recruiting, preparing and retaining high quality teachers: An empirical synthesis*. Tempe, AZ: Arizona State University, Education Policy Studies Research Unit. Retrieved May 3, 2004, from <http://www.asu.edu/educ/epsl/EPRU/documents/EPsl-0302-102-EPRU.doc>.
- Bradshaw, L.K., and Hawk, P. (1996). *Teacher certification: Does it really make a difference in student achievement?* Greenville, NC: ENCCARE (Eastern North Carolina Consortium for Assistance and Research in Education) Press.
- Brewer, T.M. (2003). "The 'Grand Paradox' in teacher preparation and certification policy." *Arts Education Policy Review*, 104(6), 3-10.
- Brophy, J. (1986, October). "Teacher influences on student achievement." *American Psychologist*, 1069-1077.
- Brophy, J., and Good, T. (1986). "Teacher behavior and student achievement." In M. Wittrock (Ed.), *Handbook on research on teaching* (3rd ed.). New York: MacMillan, 328-375.
- Bruno, J., and Doscher, M. (1981). "Contributing to the harms of racial isolation: Analysis of requests for teacher transfer in large urban school district." *Education Administration Quarterly*, 17(2), 93-108.

- Buday, M.C., and Kelly, J.A. (1996). "National Board Certification and the teaching profession's commitment to quality assurance." *Phi Delta Kappan*, 78(3), 215-230.
- Camphire, G. (2001, October). "Are our teachers good enough?" *SEDLetter*, XIII(2). Retrieved May 3, 2004, from <http://www.sedl.org/pubs/sedletter/v13n02/1.html>.
- Carlson, R.E. (1989). "An alternative methodology for appraising the job relatedness for NTE program tests." *Applied Measurement in Education*, 2(3), 243-253.
- Carnoy, M., and Loeb, S. (2002). "Does external accountability affect student outcomes? A cross-state analysis." *Educational Evaluation and Policy Analysis*, 24(4), 305-331.
- Clarridge, P.B. (1990). "Multiple perspectives on the classroom performance of certified and uncertified teachers." *Journal of Teacher Education*, 41(4), 15-25.
- Cochran-Smith, M., and Fries, M.K. (2001, November). "Sticks, stones and ideology: The discourse of reform in teacher education." *Educational Researcher*, 30(8), 3-15.
- Cohen, D.K., and Hill, H.G. (2000). "Instructional policy and classroom performance: The mathematics reform in California." *Teachers College Record*, 102(2), 294-343.
- Cooper, C.C., Benz, C.R., and Thompson, S.J. (1988). "Design for excellence in teacher education." *Journal of Teacher Education*, 39(4), 15-20.
- Curran, B., Abrahams, C., and Clarke, T. (2001, February). *Solving teacher shortages through license reciprocity*. Denver, CO: State Higher Education Executive Officers. Retrieved May 5, 2004, from <http://www.sheeo.org/quality/mobility/reciprocity.PDF>.
- Darling-Hammond, L. (1994, September). "Who will speak for the children? How 'Teach for America' hurts urban schools and students." *Phi Delta Kappan*, 76(1), 21-34.
- Darling-Hammond, L. (1997, November). *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. (1999). "Teaching and knowledge: Policy issues posed by alternative certification for teachers." *Peabody Journal of Education*, 67(3), 123-154.
- Darling-Hammond, L. (2000). *Solving the dilemma of teacher supply, demand and standards: How can we ensure a competent, caring and qualified teacher for every child?* New York: National Commission on Teaching and America's Future, Teachers College, Columbia University. Retrieved May 3, 2004, from <http://www.nctaf.org/article/?c=4andsc=17andssc=0anda=21>

- Darling-Hammond, L., and Hudson L. (1990). "Precollege science and mathematics teachers: Supply, demand and quality." In C. B. Cazden (Ed.), *Review of research in education*, 16. Washington, DC: American Educational Research Association, 223-264.
- Darling-Hammond, L., and Sykes, G. (2003). "Wanted: A national teacher supply policy for education: The right way to meet the 'Highly Qualified Teacher' challenge." *Education Policy Analysis Archives*, 11(33). Retrieved May 5, 2004, from <http://epaa.asu.edu/epaa/v11n33/>.
- Darling-Hammond, L., and Youngs, P. (2002, December). "Defining 'highly qualified teachers': What does 'scientifically-based research' actually tell us?" *Educational Researcher*, 31(9), 13-25.
- Davis, H. (2001). "The quality and impact of relationships between elementary school students and teachers." *Contemporary Educational Psychology*, 26(4), 431-453.
- Denton, J.J., and Peters, W.H. (1988, September). *Program assessment report: Curriculum evaluation of a non-traditional program for certifying teachers*. (ERIC Document reproduction service no. ED300361). College Station, TX: Texas A&M University.
- Denton, J.J., and Smith, N.L. (1984). *A cost effectiveness evaluation to alternate secondary level teacher preparation programs*. Paper presented at the annual meeting of the American Educational Research Association. New Orleans, LA.
- Dorsey, S. (1980). "The occupational licensing queue." *Journal of Human Resources*, 15(3), 424-434.
- Druva, C.A., and 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., and Zlotnik, M. (1985). "Making a difference in educational quality through teacher education." *Journal of Teacher Education*, 36(3), 2-12.
- Fenstermacher, G.D. (2002, August-September). "Reconsidering the teacher education reform debate: A commentary on Cochran-Smith and Fries." *Educational Researcher*, 31(6), 20-22.
- Fenstermacher, G.D., and Richardson, V. (2000, August). *On making determinations of quality in teaching*. A paper prepared at the request of the Board on International Comparative Studies in Education of the National Academy of Sciences.
- Ferguson, R.F., and 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*. Washington, DC: Brookings Institution, 265-298.

- Fidler, P. (2002). *The relationship between teacher instructional techniques and characteristics and student achievement in reduced size classes*. (Planning, Assessment and Research Division Publication No. 120). Los Angeles: Unified School District Program Evaluation and Research Branch.
- Finn, C.E. (1999). *The teachers we need and how to get more of them*. Washington, DC: The Thomas B. Fordham Foundation.
- Finn, C.E., Jr., and Kanstoroom, M. (2000, Summer). "Education: Solutions improving, empowering, dismantling." *The Public Interest*, 140.
- Finn, C.E., Jr., Kanstoroom, M., and Petrilli, M. *The quest for better teachers: Grading the states*. Washington, DC: The Thomas B. Fordham Foundation.
- Galambos, E.C. (1984). *Testing teachers for certification and recertification*. Washington, DC: National Commission on Excellence in Teacher Education.
- Glass, G.V. (2002). "Teacher characteristics." In A. Molnar (Ed.), *School reform proposals: The research evidence* (Chapter 8). Retrieved May 5, 2004, from <http://www.asu.edu/educ/epsl/EPRU/documents/EPRU%202002-101/epru-2002-101.htm>.
- Glass, G.V. (1990). "Using student test scores to evaluate teachers." In J. Millman and L. Darling-Hammond (Eds.), *The new handbook of teacher evaluation*. Newbury Park, CA: Sage, 229-240.
- Goldhaber, D. (2003, October). *Alternative licensure: What we know now and what we still need to learn*. Conference presentation, American Enterprise Institute for Public Policy Research, Washington, DC.
- Goldhaber, D. (2002, Spring). "The mystery of good teaching." *Education Next*. Retrieved May 5, 2004, from <http://www.educationnext.org/20021/50.html>.
- Goldhaber, D., and Anthony, E., (2003). *Indicators of teacher quality*. (ERIC Digest #184). New York: ERIC Clearinghouse on Urban Education.
- Goldhaber, D.D., and Brewer, D.J. (1999). "Teacher licensing and student achievement." In M. Kanstoroom and C.E. Finn Jr. (Eds.), *Better teachers, better schools*. Washington, DC: Thomas B. Fordham Foundation, 83-102.
- Goldhaber, D.D., and Brewer, D.J. (2001). "Evaluating the evidence on teacher certification: A rejoinder." *Educational Evaluation and Policy Analysis*, 23(1), 7.
- Good, T., and Brophy, J. (2000). *Looking in classrooms* (8th ed.). New York: Longman.

- Grossman, P.L. (1989). "Learning to teach without teacher education." *Teachers College Record*, 91(2), 191-208.
- Haberman, M. (1995, June). "Selecting 'Star' teachers for children and youth in urban poverty." *Phi Delta Kappan*, 777-781.
- Haberman, M. (1986, June). "Licensing teachers: Lessons from other professions." *Phi Delta Kappan*, 67(10), 719-722.
- Haertel, E. (1986). "The valid use of student achievement measures for teacher evaluation." *Educational Evaluation and Policy Analysis*, 8(1), 45-60.
- Hanushek, E. (1992). "The trade-off between child quantity and quality." *Journal of Political Economy*, 100, 84-117.
- Hanushek, E.A. (1994). *Making schools work: Improving performance and controlling costs* (1st Ed.). Washington, DC: The Brookings Institution.
- Hanushek, E.A. (1996a). "School resources and student performance." In G.T. Burtless (Ed.), *Does money matter? The effect of school resources on student achievement and adult success*. Washington, DC: The Brookings Institution, 43-73.
- Hanushek, E.A. (2002). "Teacher quality." In L.T. Izumi and W.M. Evers (Eds.), *Teacher quality*. San Francisco: Hoover Institution Press, 1-12.
- Hanushek, E.A., Kain, J., and Rivkin, S. (1998). *Teachers, schools and academic achievement*. (Working paper w6691). Cambridge, MA: National Bureau of Economic Research.
- Hanushek, E.A., Kain, J.F., and Rivkin, S.G. (1999, April). "Do higher salaries buy better teachers?" (NBER working paper #7082). *National Bureau of Economic Research*. Retrieved May 5, 2004, from <http://www.nber.org/papers/w7082>.
- Hanushek, E.A., and Rivkin, S.G. (2003, May). *How to improve the supply of high quality teachers*. Washington, DC: Brookings Institution.
- Hanushek, E.A., Rivkin, S.G., and Taylor, L.L. (1995). *Aggregation bias and the estimated effects of school resources*. Rochester, NY: University of Rochester, Center for Economic Research.
- Haycock, K. (1998). "Good teaching matters: How well-qualified teachers can close the gap." *Thinking K-16*, 3(2), 1-14.
- Haycock, K. (1999, March). "Good teaching matters... a lot." In K. Haycock, *Good teaching matters: How well-qualified teachers can close the gap (Results)*. Oxford, OH: National Staff Development Council.

- Hess, F.M. (2001). *Tear down this wall: The case for a radical overhaul of teacher certification*. Washington, DC: Progressive Policy Institute. Retrieved May 5, 2004, from http://www.ppionline.org/documents/teacher_certification.pdf.
- Hess, F.M. (2002, September). "Advocacy in the guise of research: The Laczko-Kerr-Berliner study of teacher certification." *21st Century Schools Project Bulletin*, 2(18). Retrieved May 5, 2004, from http://www.ppionline.org/ppi_ci.cfm?cp=1&knlgAreaID=110&subsecid=900001&contentid=250847.
- Hess, F.M. (2002). "Break the link: The fact that schools of education could no longer rely on a captive body of aspiring teachers would expose them to the cleansing winds of competition." *Education Next*, 2(1), 22-28.
- Hess, F.M. (2002). "A modest proposal." (ERIC Document Reproduction Service No. EJ642953). *American School Board Journal*, 189(4), 28-29.
- Hickok, E.W. (2002). "Teacher quality accountability systems: The view from Pennsylvania." In L.T. Izumi and W.M. Evers (Eds.), *Teacher quality*. San Francisco: Hoover Institution Press, 25-32.
- Hirsch, E., Koppich, J.E., and Knapp, M.S. (2001). *Revisiting what states are doing to improve the quality of teaching: An update on patterns and trends*. Seattle, WA: Center for the Study of Teaching and Policy, University of Washington.
- Huang, S., Yi, Y., and Haycock, K. (2002). "Interpret with caution: The first state Title II reports on the quality of teacher preparation." *The Education Trust*, 3-21. Retrieved May 5, 2004, from <http://www2.edtrust.org/NR/rdonlyres/305751B5-7635-4F5E-9BBE-068CF0849150/0/titleII.pdf>.
- Ingersoll, R. (1999). "The problem of underqualified teachers in American secondary schools." *Educational Researcher*, 28(2), 26-37.
- Ingersoll, R. (2001, Fall). "Teacher turnover and teacher shortages: an organizational analysis." *American Educational Research Journal*, 499-534.
- Jaeger, R. (1990). "Establishing standards for teacher certification tests." *Educational Measurement: Issues and Practice*, 9(4), 15-20.
- Jerald, C.D., and Ingersoll, R.M. (2002, August). *All talk, no action: Putting an end to out-of-field teaching*. Washington, DC: The Education Trust.
- Johnson, K.A. (2000, September). "The effects of advanced teacher training in education on student achievement." *Report of the Heritage Center for Data Analysis* (No. 00-09), 1-17.

- Kanstoroom, M., and Finn, C.E. (Eds.) (1999). *Better teachers, better schools*. Washington, DC: Thomas B. Fordham Foundation.
- Kaplan, L.S., and Owings, W.A. (2001, November). "Teacher quality and student achievement: Recommendations for principals." *NASSP Bulletin*, 85(628).
- Kaplan, L.S., and Owings, W.A. (2002). *Teacher quality, teaching quality and school improvement*. Bloomington, IN: Phi Delta Kappan Educational Foundation.
- Kaplan, L.S., and Owings, W.A. (2002). *Enhancing teacher quality*. Bloomington, IN: Phi Delta Kappa Educational Foundation.
- Kaplan, L.S., and Owings, W.A. (2003, May). "The politics of teacher quality." *Phi Delta Kappan*, 84(09), 687-692.
- Kearns, J. (1984). "State competency testing for teacher certification." *Journal of Teacher Education*, 35(2), 58-61.
- Kennedy, M.M. (1991). "Policy issues in teacher education." *Phi Delta Kappan*, 72, 659-665.
- Kleiner, M.M. (2000). "Occupational licensing." *Journal of Economic Perspectives*, 14(4), 189-202.
- Koppich, J., and Merseth, K. (2000). "Preparation in a five-year program." In L. Darling-Hammond (Ed.), *Studies of excellence in teacher education: Preparation at the graduate level*. New York: American Association of Colleges for Teacher Education.
- Kwiatkowski, M. (1999). "Debating alternative teacher certification: A trial by achievement." In M. Kanstoroom and C.E. Finn, Jr. (Eds.), *Better teachers, better schools*. Washington, DC: Thomas B. Fordham Foundation, 215-238.
- Laczko-Kerr, I., and Berliner, D.C. (2003, May). "In harm's way: How undercertified teachers hurt their students." *Educational Leadership*, 34-39.
- Lee, V.E., Dedrick, R.F., and Smith, J.B. (1991, July). "The effect of the social organization of schools on teachers' efficacy and satisfaction." *Sociology of Education*, 64(3), 190-208.
- Lewis, L., Parsad, B., Carey, N., Bartfai, N., Farris, E., Smerdon, B., and Greene, B. (1999). *Teacher quality: A report on the preparation and qualifications of public school teachers*. (NCES 1999-080). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.
- Lilly, M.S. (1992). "Research on teacher licensure and state approval of teacher education programs." *Teacher Education and Special Education*, 15(2).
- Lopez, O.S. (1995). "The effect of the relationship between classroom student diversity and teacher capacity on student performance: Conclusions and recommendations for

- educational policy and practice.” *The Strategic Management of the Classroom Learning Enterprise Research* series, Austin, TX.
- Mehrens, W.A., and Phillips, S.E. (1989). “Using college GPA and test scores in teacher licensure decisions: Conjunctive versus compensatory models.” *Applied Measurement in Education*, 2(4), 277-288.
- Mendro, R.L. (1998). “Student achievement and school and teacher accountability.” *Journal of Personnel Evaluation in Education*, 12, 257-267.
- Mikulecky, M., Shkodriani, G., and Wilner, A. (2004). “A growing trend to address the teacher shortage.” *ECS Policy Brief*. Denver, CO: Education Commission of the States.
- Millman, J. (1997). *Grading teachers, grading schools*. Thousand Oaks, CA: Corwin Press.
- Millman, J., and Sykes, G. (1992). *The assessment of teaching based on evidence of student learning: An analysis*. (Research Monograph No. 2). Washington, DC: National Board for Professional Teaching Standards.
- Mitchell, R., and Barth, P. (1999). “How teacher licensing tests fall short.” *Thinking K-16*, 3(1), 3-19.
- Murnane, R. (1983). “Understanding the sources of teaching competence: Skills, choices and the limits of training.” *Teachers College Record*, 84, 564-569.
- Natriello, G., and Zumwalt, K. (1993). “New teachers for urban schools? The contribution of the provisional teacher program in New Jersey.” *Education and Urban Society*, 26(1), 49-62.
- Nelson, B., and Wood, L. (1985). “The competency dilemma.” *Action in Teacher Education*, 7(1-2), 45-58.
- New York City Board of Education. (2000). *Impact of teacher certification on reading and mathematics performance in elementary and middle schools in New York City*. (Flash Report #2). New York: Author.
- Palmaffy, T. (1999). “Measuring the teacher quality problem.” In M. Kanstoroom and C. E. Finn, Jr. (Eds.), *Better teachers, better schools*. Washington, DC: Thomas B. Fordham Foundation, 19-30.
- Podgursky, M. (2001). “Defrocking the National Board: Will the imprimatur of ‘board certification’ professionalize teaching?” *Education Next*, 1(2), 79-82.
- Pratte, R., and Rury, J.L. (1991). “Teachers, professionalism and craft.” *Teachers College Record*, 93(1), 59-72.

- Ravitch, D. (2000). *Left back: A century of failed school reforms*. New York: Simon and Schuster.
- Raymond, M., and Fletcher, S. (2002, Spring). "The Teach for America evaluation." *Education Next*, 62-68.
- Raymond, M., Fletcher, S.H., and Luque, J. (2001). *Teach for America: An evaluation of teacher differences and student outcomes in Houston, TX*. Palo Alto, CA: CREDO.
- Rice, J.K. (2003). *Teacher quality: Understanding the effectiveness of teacher attributes*. Washington, DC: Economics Policy Institute.
- Rotherham, A., and Mead, S. (2003). *Back to the future: The history and politics of state teacher licensure and certification*. Washington, DC: American Enterprise Institute/Progressive Policy Institute.
- Salyer, B.A. (2003). "Alternatively and traditionally certified teachers: The same but different." *NASSP Bulletin*, 87(636), 16-27.
- Sanders, W.L., Saxton, A.M., and Horn, S.P. (1997). "The Tennessee value-added assessment system: A quantitative, outcomes-based approach to educational assessment." In J. Millman (Ed.), *Grading teachers, grading schools*. Thousand Oaks, CA: Corwin Press.
- Schaeffer, B. (1996). *Standardized tests and teacher competence*. Cambridge, MA: Fair Test.
- Shinkfield, A.J., and Stufflebeam, D.L. (1995). *Teacher evaluation: Guide to effective practice*. Kalamazoo, MI: Center for Research on Educational Accountability and Teacher Evaluation.
- Shulman, J. (1989). "Blue freeways: traveling the alternative route with big-city teacher trainees." *Journal of Teacher Education*, 40(5), 2-8.
- Smith, I.L., and Hambleton, R.K. (1990). "Content validity studies of licensing examinations." *Educational Measurement: Issues and Practice*, 9(4), 7-10.
- Soler, S. (1999). "Teacher quality is job one: Why states need to revamp teacher education." *Spectrum: The Journal of State Government*, 72(2), 23.
- Stedman, C.H. (1984). "Testing for competency: A pyrrhic victory?" *Journal of Teacher Education*, 35(2), 2-5.
- Steiner, D. (2003, October). *Preparing teachers: Are American schools of education up to the task?* Paper presented at American Enterprise Institute for Public Policy Research and the Progressive Policy Institute Conference.

- Stevens, C.J., and Dial, M. (1993). "A qualitative study of alternatively certified teachers." *Education and Urban Society*, 26(1), 63-77.
- Stone, J.E. (2002a). "The value-added achievement gains of NBPTS-certified teachers in Tennessee: A brief report."
- Stone, J.E. (2002b). "Teacher training and pedagogical methods." In L.T. Izumi and W.M. Evers (Eds.), *Teacher quality*. San Francisco: Hoover Institution Press, 33-54.
- Stronge, J.H. (2002). *Qualities of effective teachers*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Stronge, J.H., and Hindman, J.L. (2003, May). "Hiring the best teachers." *Educational Leadership*, 48-52.
- Stronge, J.H., and Tucker, P.D. (2000). *Teacher evaluation and student achievement*. Washington, DC: National Education Association.
- Stronge, J.H., and Ward, T.J. (2002, June). *Alexandria City Public Schools teacher effectiveness study*. Report presented to the Alexandria (Virginia) City School Board.
- Sullivan, C.C. (2001). *Into the classroom: Teacher preparation, licensure and recruitment*. Alexandria, VA: National School Boards Association.
- Sykes, G. (1999). "Teacher and student learning: Strengthening their connections." In L. Darling-Hammond and G. Sykes (Eds.), *Teaching as the learning profession*. San Francisco: Jossey-Bass, 151-179.
- Tatto, M.T. (2000, October). *Teacher quality and development: Empirical indicators and methodological issues in the comparative literature*. Paper commissioned by the Board on International Comparative Studies in Education of the National Academies/National Research Council.
- Teacher Education Partnership Commission (2002). *A quality teacher for every Arizona classroom*. Retrieved May 4, 2004, from <http://www.teacherpartner.com/>.
- The Education Trust (2003, September). *In need of improvement: Ten ways the U.S. Department of Education has failed to live up to its teacher quality commitments*. Washington, DC: Author.
- Trevisan, M.S. (2002, June). "The states' role in ensuring assessment competence." *Phi Delta Kappan*, 83(10), 766-771.
- Tryneski, J. (1997). *Requirements for certification of teachers, counselors, librarians, administrators for elementary and secondary schools*. Chicago, IL: University of Chicago Press.

- Walsh, K. (2002, Spring). "Positive spin: The evidence for traditional teacher certification, reexamined." *Education Next*, 2(1), 79-85.
- Walsh, K. (2001a). *Teacher certification reconsidered: Stumbling for quality*. Baltimore, MD: The Abell Foundation.
- Walsh, K. (2001b). *Teacher certification: An idea whose time has gone*. Baltimore, MD: The Abell Foundation.
- Wayne, A.J. (1999). *Teaching policy handbook: Higher passing scores on teacher licensure examination*. (Working paper). National Partnership for Excellence and Accountability in Teaching, funded by OERI.
- Weiner, L. (1990). "Preparing the brightest for urban schools." *Urban Education*, 25(3), 258-273.
- Whitehurst, G.J. (2002). "Improving teacher quality." (Symposium: Trends in Education Reform). *Council of State Governments*, 75(3), 12-15.
- Wilcox, D.D. (1999). "The National Board for Professional Teaching Standards: Can it live up to its promise?" In M. Kanstoroom and C.E. Finn Jr. (Eds.), *Better teachers, better schools*. Washington, DC: Thomas B. Fordham Foundation, 163-198.
- Wilkerson, J.R., and Lang, W.S. (2003). "Portfolios, the Pied Piper of teacher certification assessments: Legal and psychometric issues." *Education Policy Analysis Archives*, 11(45).
- Wilson, S.M., and Floden, R. (2002). *Creating effective teachers: Concise answers for hard questions*. Addendum to the report *Teacher preparation research: Current knowledge, gaps and recommendations*. Seattle, WA: Center for the Study of Teaching and Policy, University of Washington in collaboration with Michigan State University.
- Wilson, S.M., Floden, R.E., and Ferrini-Mundy, J. (2001). *Teacher preparation research: Current knowledge, gaps and recommendations*. Seattle, WA: University of Washington.
- Wilson, S.M., Floden, R.E., and Ferrini-Mundy, J. (2002). "Teacher preparation research: An insider's view from the outside." *Journal of Teacher Education*, 53(3), 190-204.
- Wong, K.K., Meyer, S.J., and Shen, F.X. (2003). *Educational resources and achievement gaps in high poverty schools: Findings from the longitudinal evaluation of school change and performance (LESCP) in Title I Schools*. Report prepared for the U.S. Department of Education under contract number EA96008001.

Wright, S.P., Horn, S.P., and Sanders, W.L. (1997). "Teacher and classroom context effects on student achievement: Implications for teacher evaluation." *Journal of Personnel Evaluation in Education*, 11(1), 57-67.

Youngs, P., Odden, A., and Porter, A.C. (2003). "State policy related to teacher licensure." *Educational Policy*, 17(2), 217-237.

Zumwalt, K. (1991). "Alternative routes to teaching: Three alternative approaches." *Journal of Teacher Education*, 42(2), 83-92.

Zumwalt, K. (1996). "Simple answers: Alternative teacher certification." *Educational Researcher*, 25(8), 40-42.

GLOSSARY

analysis of variance – statistical analysis by which sources of variability can be identified

aggregated data – data for which individual scores on a measure have been combined into a single group summary score

Example:

In education research, it is common to aggregate individual student scores on an achievement test into a mean score for each school. Researchers then use the aggregate school achievement score for data analyses. Aggregating data reduces the sample size (e.g., from 5,000 students to 10 schools.) Aggregating data also obscures differences among individual scores.

bias – any effect that is introduced into an experiment or research study that may influence the outcome based on anything other than the variables involved (e.g. expectations, the use of inappropriate statistics)

comparative descriptive study – a research study in which data are collected to describe and compare two or more groups of participants or entities

Example:

A researcher identifies high-poverty schools in the state that have either high or low student achievement. The researcher describes the alignment or match between each school's curriculum and state standard, and compares the high- versus low-achieving schools to determine whether the degree of alignment is different.

concurrent validity – a method of establishing validity for a given assessment instrument by comparing the outcome or findings of the instrument being investigated with the outcome or findings of an assessment instrument already established as valid

control – the strategy used in scientific research to regulate the effects of variables that are not intended to influence the results or conclusions

Example:

A researcher conducts a study of two different teacher preparation courses on how to teach mathematics. The researcher controls for differences among preservice students by randomly assigning the students to one of the two courses. The researcher controls for differences among course instructors by having a single instructor teach both courses.

construct validity – whether an assessment tool measures the construct, or some attribute or quality which is not “operationally defined” that it purports to measure

content validity – the extent to which an assessment tool represent all aspects of the construct it is intended to measure

correlational research/study – nonexperimental research in which data are collected to determine the relationship between them

Example:

In School District X, a researcher collects data on beginning teachers' scores on the state licensing test (variable 1) and data on the achievement gains of each teacher's students (variable 2). The researcher then uses correlational statistics to measure the association between the two variables.

cross-sectional data – data gathered from a cross-section of the sample of interest that is assumed to represent the population as it moves through the stages measured

dependent variable – the variable measured in a study – the “outcome”; in experimental research, the dependent variable is affected by the independent variable; in correlational research, the dependent variable is associated with one or more other variables

Examples:

In an experimental research study, a researcher randomly assigns teachers in a large elementary school to receive one of three types of professional development: a class on instructional strategies, a training program on how to increase student motivation or a teacher discussion group. The researcher measures the differences in achievement gains among the students of the three teachers. The dependent variable is student achievement gains.

For a correlational research study, a researcher collects data on beginning teachers' scores on the state licensing test and data on the achievement gains of each teacher's students (variable 2). The researcher then uses the association between the two variables to estimate student achievement gains. The dependent variable is student achievement gains.

econometric method – an economic model that describes and tests economic relationships to obtain a measure of the strengths of the influences of the different variables

effect size – the degree to which a practice, program or policy has an effect based on research results, measured in units of standard deviation

Example:

A researcher finds an effect size of $d = .5$ for the effect of an after-school tutoring program on reading achievement. This means (provided that the research study is valid) that the average student who participates in the tutoring program will achieve one-half standard deviation above the average student who does not participate. If the standard deviation is eight points, then the effect size translates into four additional points, which will increase a student's ranking on the test.

empirical research/empirical studies – research that seeks systematic information about something that can be observed in the real world or in a laboratory

experimental study (experimental research) – a research study that has the goal of determining whether something causes an effect

external validity – the degree to which results from a study can be generalized to other participants, settings, treatments and measures

hierarchical linear modeling (HLM) – a statistical technique used to analyze data that are from participants who exist within different levels of a hierarchical structure. For example, student achievement data reflect influences from the family, classroom, grade, school, district and state. Through HLM, the influences of these different levels on student achievement can be estimated.

hypothesis(es) – a statement about the researcher’s expectations concerning the results of a study

Examples:

A new standards-based mathematics curriculum will benefit elementary students at all grade levels.

A new standards-based mathematics curriculum will have different effects on elementary students depending on grade level.

independent variable – in experimental research, the variable that the researcher varies or manipulates to determine whether it has an effect on the dependent variable

Example:

As part of an experiment, a researcher randomly assigns teachers in a large elementary school to receive one of three types of professional development: a class on instructional strategies, a training program on how to increase student motivation or a teacher discussion group. The researcher measures the differences in achievement gains among the students of the three teachers. The independent variable is professional development, and it has three different values.

interrater reliability – the degree to which multiple raters of a non-objective assessment procedure agree as to its rating on a given scale

longitudinal data – data collected from the same participants at different points in time; the purpose is to make conclusions about individual change over time

Example:

A researcher studies the mathematics achievement of students who were taught a new standards-based mathematics curriculum when they were in 6th grade. The researcher compares their performances in mathematics achievement in grades 7, 8 and 9 to the performance of another group of students at each of those grade levels who were not taught the new curriculum in 6th grade. The purpose of the research is to determine

whether change in mathematics performance over time is related to the type of 6th-grade mathematics curriculum.

meta-analysis – a comprehensive, systematic and quantitative review of past empirical research studies on a specific topic; most meta-analyses examine only quantitative studies; effect-size statistics are calculated to produce an overall conclusion about the various studies on the topic

Example:

A researcher conducts a meta-analysis of computer-assisted instruction in reading. The researcher examines 40 studies and calculates an overall effect size of $d = .25$, indicating a small positive effect of computer-assisted instruction on reading achievement.

multiple regression analysis – a statistical technique that determines the linear association between a set of predictor variables and a dependent variable, and identifies the combination of predictor variables that best estimate the dependent variable

Example:

In School District X, a researcher collects data on beginning teachers' scores on the state licensing test (predictor 1), number of college courses in mathematics (predictor 2), amount of time spent in school-based field experiences prior to certification (predictor 3) and the achievement gains in mathematics by each teacher's students (dependent or criterion variable). The researcher uses multiple-regression statistics to measure the association between the three teacher variables and student achievement gains and to estimate student achievement gains based on the contribution of each of the teacher variables to that association.

negative correlation – a relationship between two variables in which large values of one variable are associated with small values of the other

peer-reviewed – a research study that has been critiqued by other researchers prior to publication or presentation at a research conference

practical significance – the degree to which a practice, program or policy has enough of an effect to justify its adoption

production function analysis – an analysis by which an input measure is related to an output measure using a statistical technique such as correlation or multivariate analysis (regression analysis)

proxy – a measure used to approximate the data sought when it is difficult to get a more precise measure due to constraints involving data collection or time

Example:

Passing rates on state licensing tests by teacher candidates are a proxy measure for the quality of teacher preparation delivered by teacher education institutions.

psychometric – the field of study concerned with the measurement of psychological aspects of a person such as knowledge, skills or abilities

qualitative research – research in which the data are narrative descriptions or observations

Example:

A researcher observes how teachers instruct different reading curricula in two different schools. The researcher also interviews the teachers to understand their approaches to the different curricula and how approaches might be influenced by school characteristics.

quantitative research – research in which the data are numbers and measurements

Example:

A researcher randomly assigns students to different reading curricula. At the end of the school year, the researcher examines the students' scores on a reading achievement test to determine whether the different curricula had different effects on reading.

quasi-experimental study – a research study in which (1) an independent variable is directly manipulated to measure its effects on a dependent variable and (2) participants are *not randomly assigned* to comparison groups

Example:

A researcher assigns 15 teacher preparation candidates who have senior seminar on Wednesdays to participate in eight weeks of student teaching. The researcher assigns 15 teacher preparation candidates who have senior seminar on Tuesdays to participate in 16 weeks of student teaching. After the candidates graduate, the researcher compares their scores on a performance-based teacher-licensing test. The amount of student teaching is the independent variable, and candidate performance on the teacher-licensing test is the dependent variable. The researcher does not randomly assign candidates to the comparison groups. As a result, differences between the groups on the test could be due to the amount of student teaching or due to other characteristics of the teacher candidates.

regression analysis – a statistical technique for determining the association between a dependent variable and one or more independent variables and thereby being able to predict variation in dependent variable by knowing the other variables

Example:

In School District X, a researcher collects data on beginning teachers' scores on the state licensing test (variable 1), number of college courses in mathematics (variable 2), amount of time spent in school-based field experiences prior to certification (variable 3) and the achievement gains in mathematics by each teacher's students (dependent variable). The researcher uses regression statistics to measure the association between the three teacher variables and student achievement gains and to estimate student

achievement gains based on the contribution of each of the teacher variables to that association.

reliability – the extent to which an assessment instrument yields consistent results over repeated observations or measurements

replicate – to repeat a research study using the same method and similar participants; a successful replication obtains the same results as the original study

sample size – the number of participants (e.g., students) or entities (e.g., schools) in a study sample; large samples are preferred because, if randomly selected, they are more representative of the population than small samples

selection bias – systematic effects on the dependent variable that occur due to characteristics of the study participants

Example:

A researcher conducts a study on the influence of student teaching on teaching performance. The researcher assigns 20 teacher preparation candidates who attend college during the day to participate in 16 weeks of student teaching. The researcher assigns 20 candidates who are night students to have eight weeks of student teaching. Selection bias in this study is likely because the characteristics of day and night students, such as age and motivation, may be different. The results could be due to these differences instead of the amount of student teaching.

simple descriptive study – a research study in which data are collected to describe persons, organizations, settings or phenomena

Example:

A researcher surveys administrators of 10 alternative teacher preparation programs to describe characteristics of the different programs.

standard deviation – a measure of the variability of the scores in a distribution (i.e., a set of scores) equivalent to the average distance of the scores from the mean

Example:

Scores: 9, 10, 10, 12, 14

For the example set of five scores, the mean is 11, and the standard deviation is 2. The scores vary on average about two points from the mean.

statistical significance – a result that has 5% or less probability of occurring by chance; because it is unlikely that a statistically significant result has occurred by chance, the result is said to reflect non-chance factors in the study, such as the effects of a treatment

stratified purposive sampling – a method of sampling a population of interest with a purpose in mind; in other words, certain levels of a population are designated specifically based on the hypothesis or study; the counterpoint to this would be random sampling

structural equation modeling (SEM) – a statistical method generally used for confirmatory rather than exploratory purposes, to determine the extent to which data on a set of variables are consistent with hypotheses about the association among the variables

synthesis(es) – a comprehensive and systematic literature review of past empirical research studies on a specific topic; research syntheses can be quantitative or qualitative; meta-analysis is the term used for a quantitative synthesis, and narrative review is the term used for a qualitative synthesis

validity – the extent to which a study or measure accurately reflects or assesses the specific concept or variable the researcher is attempting to measure

variable – a characteristic or quantity that can change and have different values

Example:

Variables studied in education include characteristics of students (e.g., achievement), teachers (e.g., certification), schools (e.g., curriculum), districts (e.g., leadership), teacher preparation programs (e.g., accreditation) and states (e.g., education funding).