SYSTEMATIC EVALUATION FOR CONTINUOUS IMPROVEMENT OF TEACHER PREPARATION

Volume 1: Cross-Case Analysis

Regional Educational Laboratory
Contract #ED-01-CO-0006

Prepared by
Ceri B. Dean, Ph.D.
Patricia A. Lauer, Ph.D.

Mid-continent Research for Education and Learning
2550 S. Parker Road, Suite 500
Aurora, CO 80014
303.337.0990 (phone)
303.337.3005 (fax)
www.mcrel.org

Prepared for
Institute of Education Sciences
U.S. Department of Education
Washington, D.C. 20208

December 2003

Mid-continent Research for Education and Learning
2550 S. Parker Road, Suite 500
Aurora, CO 80014
303.337.0990
303.337.3005 (fax)
www.mcrel.org

This report has been funded at least in part with federal funds from the U.S. Department of Education under contract number ED-01-CO-0006. The content of this publication does not necessarily reflect the views or policies of the U.S. Department of Education nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

TABLE OF CONTENTS

Introduction 1

Background and Rationale 3

Accountability for Teacher Education 4

Standards for Teacher Education 5

Evaluation and Assessment in Teacher Preparation Programs 6

Institutions Studied 9

Study Method 10

Results of Cross-Case Analysis 12

Guiding Question #1: How Are Individuals, Groups, and the Components of the Teacher Preparation Programs Evaluated? 13

Guiding Question #2: How Do Teacher Preparation Programs Align Evaluation with Program Standards/Goals? 17

Guiding Question #3: How Do Teacher Preparation Programs Develop Systematic Evaluation? 18

Guiding Question #4: How Do P–16 Stakeholders Influence Evaluation of the Teacher Preparation Programs? 20

Guiding Question #5: How Do External Influences Affect Evaluation of the Teacher Preparation Programs? 22

Guiding Question #6: What Are the Characteristics of a Culture that Supports Data Collection and Use for the Evaluation of the Teacher Preparation Programs? 22

Summary of Structures and Processes — Answering the Overall Question 25

Areas for Further Study 32

Determining Graduates' Teaching Effectiveness 32

Managing Data 33

Determining Proficiency 33

References 35

ACKNOWLEDGEMENTS

This report is volume 1 of a two-volume report resulting from a collaborative effort of Mid-continent Research for Education and Learning (McREL) and five other regional educational laboratories[1]. The authors acknowledge and greatly appreciate the many ways in which the following colleagues from those laboratories assisted with the project: Tom Barlow (PREL), Steve Bingham (SERVE), Jane Hange (AEL), Stephen Marble (SEDL), and Aida Walqui (WestEd). Sharon Horn of the U.S. Department of Education has our special thanks for advocating for the project and providing moral support. We are grateful for the assistance of our contacts at each of the sites: Kathy Lake, Alverno College; Ann Bullock, East Carolina University; Joanna Uhry, Fordham University; and Ruth Ash, Samford University. We appreciate the willingness of program faculty, administrators, and graduates, and P–12 teachers and principals to share their perspectives and knowledge of the programs with us. The authors thank Zoe Barley and Lou Cicchinelli of McREL and other individuals who reviewed this document for their helpful comments. Special thanks go to Barb Gaddy of McREL for her editorial assistance and commitment to quality.

[1] AEL, Pacific Resources for Education and Education and Learning (PREL), Southwest Educational Development Laboratory (SEDL), SERVE, and WestEd

INTRODUCTION

Good teachers make a difference. That was the message from the National Commission on Teaching and America's Future in 1996, and that is the message today in the No Child Left Behind Act of 2001, which reauthorized the Elementary and Secondary Education Act. In its call to "leave no child behind," the Act puts teacher quality center stage and calls for all teachers in core academic subject areas to be "highly qualified" by the end of the 2005–2006 school year. As defined by the Act, being highly qualified means having at least a bachelor's degree and demonstrating competency in subject knowledge and teaching skill. Without doubt, high-quality teacher preparation is more important today than ever.

Although several organizations have developed standards for various aspects of teacher preparation (e.g., the National Council for Accreditation of Teacher Education [NCATE] and the Interstate New Teacher Assessment and Support Consortium [INTASC]), there is a lack of agreement about all that constitutes high-quality teacher preparation. Consistent with its focus on teacher quality, in the late 1990s representatives from the U.S. Department of Education hosted and attended meetings that brought together teacher educators and associations concerned with teacher education (e.g., NCATE, the Teacher Education Accreditation Council [TEAC], the American Association of Colleges for Teacher Education [AACTE]) to discuss this issue and to find ways to deepen the discussion. The National Awards Program for Effective Teacher Preparation, established in 2000 by the U.S. Department of Education, grew out of these discussions. The award was designed to recognize teacher preparation programs that could present compelling evidence that their programs were effective in preparing teachers who could help all students meet high academic standards (U.S. Department of Education, 2000b). Winning programs were required to provide three types of evidence to demonstrate their effectiveness:

- Formative: Evidence that the program gathers and uses data to make adjustments to the various stages of the program (e.g., admissions, course development, field experiences, assessment of knowledge and skills)
- Summative: Evidence of the effectiveness of the overall program in helping graduates acquire the knowledge and skills needed to improve all students' learning (i.e., content knowledge, pedagogical knowledge and skills, and skills to examine beliefs about learners and teaching as a profession)
- Confirming: Evidence of the effectiveness of program graduates in K–12 settings

Applicants for the award were self-nominated and, to support the mathematics and reading priorities of the Secretary of Education at that time, were restricted to elementary

programs and secondary programs that prepared teachers of mathematics. Applications were initially reviewed by the U.S. Department of Education to determine if applicants met eligibility requirements. Next, a group of representatives from higher education that included content experts and teacher educators reviewed the applications using a set of criteria that focused on the rigor, sufficiency, and coherence of the evidence provided by the applicant. This group made recommendations about the applicants that should go forward as semi-finalists in the review process. The semi-finalists received a site visit by a team that included a regional educational laboratory staff member familiar with the awards program, a higher education representative, and a K–12 representative (in most cases a principal or superintendent). A panel of experts representing higher education, reading and mathematics content, and program evaluation reviewed the original applications and site visit summary reports prepared by the site visit teams and recommended awardees to the U.S. Department of Education. The Department made the final decision based on the recommendations of the expert panel and associated documentation.

Since one of the purposes of the awards program was to deepen discussion of high-quality teacher preparation, the U.S. Department of Education contracted with five of the regional educational laboratories to conduct a study of the structures and processes used by the four recipients of the award — Alverno College, Elementary Education Program; East Carolina University, Middle Grades Mathematics Program; Fordham University Graduate School of Education, Initial Teacher Education Elementary Program; and Samford University, Elementary Education Program — to systematically evaluate their teacher preparation programs. Systematic evaluation was selected as the focus for the study because the U.S. Department of Education was interested in furthering the discussion about how to determine the overall effectiveness of teacher preparation programs as well as the effectiveness of various program components. This report is volume 1 of a two-volume report documenting the findings of that study.

In this report, *systematic evaluation* means the formal determination of the quality and effectiveness of the teacher preparation program. At its core, systematic evaluation of teacher preparation involves the processes of inquiry and judgment. This means collecting relevant information about student and program performance through a variety of assessments and using that information to judge how well the program meets the standards established for it (Worthen & Sanders, 1987). Steps in evaluation include establishing leadership, determining a purpose, selecting information sources, developing or choosing instruments, establishing a schedule, collecting data, analyzing results, reporting and disseminating results, and implementing follow-up actions (Brainard, 1996).

This report highlights commonalities and differences in how four teacher preparation programs address systematic evaluation. The companion document to this report,

Systematic Evaluation for Continuous Improvement of Teacher Education: Volume 2: Case Summaries (Lauer & Dean, 2003), presents more detailed information about systematic evaluation at each of the individual programs. Although the case studies include some information about features of each program, such as course sequences, field placements, and assessments, the primary focus is on the structures and processes the individual sites used for systematic evaluation. The case studies do not provide detailed descriptions of program elements. This reflects the emphasis of the awards program in its inaugural year — to recognize programs that had compelling evidence of effectiveness and begin to document the types of evidence programs are able to provide around various aspects of their program. The long-term goal was to examine the programs of several years' award winners to determine which program elements contribute most to program effectiveness.

To facilitate understanding of differing roles in the four sites, a common set of terms is used throughout this report. *Candidate* refers to the preservice student who is enrolled in the teacher education program. The *field supervisor* is the person who the university employs to supervise field placements of candidates; some field supervisors also are school employees, such as principals. The *cooperating teacher* is the teacher of the classroom where candidates do their student teaching. Explanations for other terms that are specific to sites are included in the case descriptions.

The audience for this report includes teacher education researchers and other faculty and administrators involved with teacher preparation programs. This report can help teacher preparation programs address calls for accountability found in Title II of the Higher Education Act and from the field (Cochran-Smith, 2001; Howey & Zimpher, 1999; Murray, 2000; Wilson, Floden, & Ferrini-Mundy, 2001). Readers can benefit from the study by comparing the characteristics of the evaluation structures and processes of their own programs with those of the nationally recognized programs that were the focus of this study.

BACKGROUND AND RATIONALE

The quality of teacher preparation programs has been a focus of concern for a number of years (Holmes Group, 1986; National Commission on Teaching and America's Future, 1996; Tom, 1997; Darling-Hammond, 1999). Concerns have increased considerably since accountability pressures have mounted in the K–12 system and some key studies were published that linked the quality of teachers with student achievement (e.g., Sanders & Rivers, 1996; Jordan, Mendro, & Weerasinghe, 1997). Although some studies (e.g., Darling-Hammond, 2000) have been conducted that describe the content, processes, and outcomes of programs deemed exemplary by K–12 and higher education researchers, expert practitioners, and program graduates, many policymakers continue to hold the belief that there are few, if any, worthwhile teacher education programs. One intention of

the National Awards Program for Effective Teacher Preparation was to gather information about programs that could back up their claims of effectiveness with a variety of data. As described in the following sections, the rationale and background for this study are found in the increasing calls for accountability for teacher preparation, discussions about standards for teacher preparation (which are closely linked to accountability), and efforts to improve teacher preparation through program evaluation.

Accountability for Teacher Education

There is widespread agreement that teacher education programs and institutions of higher education should be held to a higher level of accountability for the performance of teacher graduates. For example, a 2000 U.S. Department of Education publication, *Eliminating the Barriers to Improving Teaching* discusses ways to eliminate barriers to improving the quality of teaching. Among the barriers cited is the lack of accountability for high-quality teacher preparation by both teacher education programs and the institutions of higher education that house them. The report calls for developing new measures of effectiveness of teacher preparation and reporting results on these measures to the public. Changes in Title II of the Higher Education Act that require evidence about the performance of teacher graduates also have heightened the national interest in teacher accountability (Wilson et al., 2001) and have made accountability a pressing issue for teacher education programs.

> This emphasis on accountability can be seen in policies associated with reform of teacher preparation. As Cochran-Smith (2001) notes, current reforms are driven by the "outcomes questions" — "What should the outcomes of teacher education be for teacher learning, professional practice, and student learning?" and "How, by whom, and for what purposes should these outcomes be documented, demonstrated, and/or measured?"

As we enter the twenty-first century, the outcomes, consequences, and results of teacher education have become critical topics in nearly all of the state and national policy debates about teacher preparation and licensure as well as in the development of many of the privately and publicly funded research agendas related to student learning. (p. 6)

Cochran-Smith (2001) identifies three sources for outcomes of teacher preparation programs: (1) long-term, general impacts, including K–12 student achievement; (2) candidates' scores on teacher tests, aggregated by programs from which they graduate; and (3) documentation of professional performances by teacher candidates. As the author notes, each of these sources for outcomes has associated controversies and implications for data collection and interpretation. She identifies two approaches to education policy — the market driven (or deregulation) approach and the democracy-driven (or professionalization of teaching) approach — and emphasizes that the struggle between these two approaches is played out in the policy arena, which complicates teacher

education programs' efforts to determine their effectiveness. For example, the market-driven approach emphasizes standardized teacher subject-matter knowledge tests to determine teacher competency, while the democracy-driven approach includes performance assessments of teacher knowledge and skills during the various phases of the preparation program.

Like Cochran-Smith, Howey and Zimpher (1999) recommend that teacher candidates be assessed throughout their preservice years to measure the nature of their development as teachers. They also stress the need to link teacher preparation and performance with K–12 student learning: "The emphasis in assessment must be squarely on coupling teacher performance and teacher learning with *pupil learning* [italics in original]" (p. 301). Thus, teacher education institutions should require their candidates to demonstrate the teaching skills they learned as a result of their preparation. According to the authors, accountability in teacher preparation is a considerable problem; the lack of evidence about the quality of graduates from teacher preparation programs casts doubt on teacher education as a valuable enterprise. There have been some efforts to address this problem, however. For example, the teaching standards of the National Council of Accreditation of Teacher Education (NCATE) involve performance-based accreditation in which education schools are judged by the performances of their candidates' K–12 students. Questions in this study were designed to illuminate how the selected programs gather data about their performance in relation to the various outcomes of teacher education.

Standards for Teacher Education

As is evident in the previous section, discussions about the quality of teacher preparation cannot be separated easily from discussions about standards. For example, the National Commission on Teaching and America's Future argues in its 1996 publication *What Matters Most: Teaching For America's Future* that "the critical issue for improving the caliber of teaching is creating a viable system for using standards to guide teacher learning and create accountability" (p. 31). The connection also is seen in Yinger's (1999) discussion of the policy background of standards. Yinger describes a shift in emphasis from teacher education inputs, such as courses, to outcomes that are measured by performance-based assessments of teacher candidates. One result of this shift was the development of standards by the National Council of Accreditation of Teacher Education (NCATE), the National Board for Professional Teaching Standards, and the Interstate New Teacher Assessment and Support Consortium. Although each organization serves different purposes, Yinger contends that their work has resulted in a consensus on effective teaching:

> Through the work of the three organizations . . . a powerful consensus has emerged regarding the definition and assessment of good teaching throughout a career, from preservice education to advanced professional certification. The

> standards have framed an image of the professional teacher as a knowledgeable, reflective practitioner willing and able to engage in collaborative, contextually grounded learning activities. (pp. 103−103)

Yinger and Hendricks-Lee (2000) explain the importance of standards for the professionalization of teaching. According to the authors, standards operate as controls for ensuring the quality of professionals whose practices are judged against an established knowledge base. Standards also "define effective practice in terms of desired outcomes" (p. 97). The authors caution, however, that to be useful in these ways, standards must guide all aspects of the teacher preparation program, including expectations for both students and teacher education faculty and the use of assessments that provide feedback for improvement in relation to the standards.

Given the importance of standards noted by Yinger and Hendricks-Lee (2000), several questions in this study focus on the link between standards and program evaluation. As Cochran-Smith (2001) indicates, however, some educators have concerns about the use of standards for accountability and improvement of teacher preparation programs. For example, the Fordham Foundation (1999) is opposed to the regulation of teacher education through accreditation standards because such regulations may serve as barriers to those who want to enter teaching through alternative paths. Murray (2000) expresses another concern in suggesting that standards such as NCATE's are based on consensus and still need validation through research on effective teaching practices. Murray supports the work of the Teacher Education Accreditation Council (TEAC), which emphasizes quality assurance systems that produce evidence about the effectiveness of teacher preparation programs and demonstrate whether graduates have "in fact acquired the knowledge, disposition, and skill their academic degree indicates" (p. 47). Murray and the TEAC are not opposed to standards but, rather, emphasize the need for standards that can be supported by evidence. In summary, although there is some disagreement about the role of standards in accountability for teacher preparation, there is wide endorsement of *accountability* for teacher preparation through the collection of evidence of effectiveness. This study includes questions that will shed light on how programs wrestle with the issue of standards' role in accountability.

Evaluation and Assessment in Teacher Preparation Programs

Prior to the 1970s, there was little interest in studying evaluation of teacher preparation programs. Interest grew, however, during the 1970s and 1980s, due in part to requirements of NCATE standards. At that time, a number of individuals and institutions fostered development of the Teacher Education Program Evaluation and Follow-Up (TEPFU) movement (Hall, 1981). Members of this group and others gathered at the 1981 AACTE meeting to summarize the state of teacher education program evaluation at the time and to make recommendations for its future. For example, Vaughn (1981) predicted

"public dissatisfaction and professional obligation will only *increase* the importance of teacher evaluation in the future" [italics in original] (p. 105). He also noted that the field of teacher education evaluation was characterized by "reliance on superficial techniques that are minimal at best, a dearth of more comprehensive techniques (e.g., direct classroom observations), and very little financial commitment to evaluation" (p. 105). He recommended that teacher education program evaluation include data on the inputs and processes of the entire spectrum of the teacher education program — throughout the undergraduate program and after graduation — and that it "be formed by more comprehensive, detailed and objective data collection; collaboration on design, implementation and use; concern with both program and products; adequate financial support; and a longitudinal perspective" (p. 112).

In the same report, Ayers (1981) provides the following design characteristics for teacher education program evaluation:

1. The faculty, administration and governing body of the institution must be committed to conducting a program of teacher evaluation.
2. A teacher evaluation program must be viewed as a part of the total teacher education program and not as an isolated project.
3. A teacher evaluation program must be specific, yet reflect the total teacher education effort. The evaluation efforts must reflect the goals and objectives of specific programs and the overall program for the training of teachers.
4. An evaluation program must be continuous and longitudinal in nature.
5. An evaluation program must provide for and reflect a knowledge of the state-of-the-art in educational research and evaluation as it relates to teacher education.
6. An evaluation program must reflect thorough planning prior to implementation and consider the cooperation of K–12 schools, selection of subjects, data sources, qualitative and quantitative data collection, points of data collection, and data collection instrumentation, and personnel training.
7. An evaluation program must reflect the most efficient use of resources for data processing, storage, retrieval, and analyses.
8. An evaluation program must provide for an effective communication system for input into the evaluation process and feedback of evaluation information.
9. An evaluation program must provide for a workable system of evaluation data utilization in program decision-making.
10. An evaluation program must have a component to assess the effectiveness of the evaluation process.

Despite the calls for advances in the field of teacher education evaluation, by the late 1980s few changes had been made. According to Peterson (1989), teacher education evaluation was dominated by single data sources such as graduate opinion follow-up surveys, ratings by principals of first-year teachers, and graduate employment rates. It was rare that programs addressed the impact of graduates on their students, schools or the profession. The prevailing attitude was that it is very difficult to document the effects of preservice programs once graduates are in the field. Peterson notes that supervisor ratings also were seen as having limited reliability due to a limited number of observations, supervisors' personal issues, and instability of teacher performance during the first year of teaching. At the time, there were increased calls for standardized tests for teachers, which Peterson notes are not useful for individual or program improvement because they do not address program components such as field experience. Instead, Peterson suggests a dossier that contains multiple sources of data related to the context in which the teacher teaches. Examples include peer review of materials, student reports, parent surveys, teacher tests, documentation of professionalism, student achievement data, systematic observation, and administrator reports. The dossier is evaluated using the elements of quality practice as defined by the institution.

Recently, there has been renewed interest in teacher education evaluation. For example, *Changing the Practice of Teacher Education: Standards and Assessment as a Lever for Change* (Diez, 1998) provides information on how teacher education programs can gather evidence of effectiveness and on the benefits of this process. This book describes the stories of several teacher preparation programs that undertook reform by

> clarifying the outcomes of their programs, developing performance assessment processes to develop and document the development of student learning outcomes, developing strategies to involve faculty across the institution and in P–12 schools in the reform effort, and designing an evaluation plan to guide continuous improvement efforts. (pp. 2–3)

Each institution met with various challenges and varying degrees of success. Across the institutions, key practices were identified that supported and sustained the change process, including the process of evaluating program effectiveness. One important focus was relationships — the institutions developed internal and external relationships that facilitated their reform efforts, and faculty found new ways of working together (Hass & Stoffels, 1998). In addition, processes that supported change were developed; among them, importantly, was the adoption of a continuous improvement model in which program evaluation was viewed as ongoing, not just an activity to be completed once and for all (Henn-Reinke & Kies, 1998).

Another key change was increased understanding that both teaching and learning are improved by the feedback about progress toward goals that performance assessment of teacher candidates provides (Diez, 1998). A study of seven teacher education programs

that the National Commission on Teaching and America's Future (NCTAF) identified as preparing teachers for successful instruction of diverse learners also found that an emphasis on the use of performance assessments was a feature common to the programs (Darling-Hammond, 2000).

Despite calls for higher education institutions to "mandate a campus-wide review of the quality of their institutions' teacher education programs" (American Council on Education, 1999) and acknowledged benefits of program evaluation (Murray, 2000), evaluation of teacher preparation is often limited in scope and done primarily to satisfy accreditation requirements (Thomas & Loadman, 2001). Further, even if institutions collect and analyze such data, they often fail to use the data to make program improvements. Thomas and Loadman (2001) call on teacher education programs to collect quantitative and qualitative data on their graduates' teaching knowledge and skills and to develop scenarios that describe characteristics of programs that are effective and positively affect teaching and teacher education. This study explores the extent to which the selected programs answer Thomas and Loadman's call and adds to the limited knowledge base about how teacher education institutions collect, analyze, and use data to monitor and improve the effectiveness of their programs.

INSTITUTIONS STUDIED

In 2000, the U.S. Department of Education recognized teacher preparation programs at four institutions through the National Awards Program for Effective Teacher Preparation. These programs were recognized as effective in preparing teachers who help improve the learning of all of their K–12 students. The following four programs are the focus of this report:

Alverno College, Elementary Education Program: Alverno College is an independent, Catholic, liberal arts college for women, located in Milwaukee, Wisconsin. For the 2001–2002 school year, enrollment for undergraduate and graduate programs was about 1,950 students. The faculty of Alverno have defined a curriculum that is organized around eight abilities that students must demonstrate at an "advanced" level in order to graduate: communication, analysis, problem solving, decision making, social interaction, global perspective, effective citizenship, and aesthetic responsiveness. One important consequence of the ability-based curriculum is the use of performance-based assessments to determine candidates' progress.

East Carolina University, Middle Grades Mathematics Program: East Carolina University is a public university located in Greenville, North Carolina, with an enrollment of approximately 19,000 students for the 2001–2002 school year. The Middle Grades Mathematics Program is one of 26 teacher education programs at the university. Faculty in the program are driven by the belief that it is impossible to be a successful middle school mathematics teacher without being a successful middle school teacher.

Candidates in the program take six courses specific to middle school preparation and eight mathematics courses, six of which were developed specifically for prospective middle grades mathematics teachers.

Fordham University Graduate School of Education, Initial Teacher Education Elementary Program: Fordham University is a private, Catholic Jesuit institution located in New York City. About 2,500 of the university's 14,600 students were enrolled in the Graduate School of Education for 2001–2002. The winning program is part of the Division of Curriculum and Teaching, which emphasizes the study of contemporary education issues and the preparation of educators based on a reflective practitioner model. Cohorts of candidates progress through a series of coordinated courses and field experiences that link theory with practice.

Samford University, Teacher Education Department: Samford University is the largest private institution of higher education in Alabama. Located in Birmingham, it has a combined enrollment of about 4,380 students in its undergraduate and graduate programs for 2001–2002. In recent years, university programs have been influenced by principles of total quality management and problem-based learning. Samford teacher education programs are based on a model of reflective decision making.

Study Method

This study was undertaken by a team of researchers representing five of the regional educational laboratories at the request of the Office of Educational Research and Improvement (now known as the Institute of Education Sciences). The purpose of the study was to support the National Awards Program for Effective Teacher Preparation by helping others understand how these programs gather credible evidence of their effectiveness in preparing teachers who can help all K–12 students meet high learning standards. By studying the structures and processes for program evaluation used by the four winners of the National Award for Effective Teacher Preparation, this study adds to the small body of research on how teacher education programs gather evidence of their effectiveness (e.g., Zimpher & Loadman, 1986; Diez, 1998; Darling-Hammond, 2000; Harris, Salzman, Frantz, Newsome, & Martin, 2000).

The following overall research question focused the study:

> What are the structures and processes of systematic evaluation that supports effective teacher preparation?

In addition, the following six guiding questions were designed to elicit details about the structures and processes of the evaluation system at each institution studied. These guiding questions formed the basis of the structured interview protocol (see Appendix A) that was used across the four sites. These questions reflect literature on the use of

evaluation for improvement and accountability of teacher preparation (e.g., Collias, 2000; Darling-Hammond, 1999, 2000; Diez, 1998; Harris et al., 2000).

1. How are individuals, groups, and the components of the teacher preparation program (i.e., liberal arts courses, clinical experiences, professional courses) evaluated?
2. How do teacher preparation programs align evaluation with program standards/goals?
3. How do teacher preparation programs develop systematic evaluation?
4. How do P–16 stakeholders influence evaluation of teacher preparation programs?
5. How do external influences (e.g., state certification requirements, national organizations) affect evaluation of teacher preparation programs?
6. What are the characteristics of a culture that supports data collection and its use for the evaluation of teacher preparation programs?

Researchers received training in data collection procedures prior to the site visits. The training included a discussion of the purpose of the site visit; review of the research design, interview protocol, and the document review protocol; a discussion of how data would be analyzed; and clarifications and answers to individual questions.

Key program participants at each site were interviewed, including the dean of the School of Education and other program administrators, Arts and Sciences faculty, teacher education faculty, K–12 principals and cooperating teachers, and program graduates. A different two-member research team conducted the interviews over the course of two days at each program site. Interviews were conducted individually or in small groups that generally consisted of two individuals, though one interview included 10 individuals. Interviews lasted one to two hours. An interview matrix (see Appendix B) identifies which questions were asked of different participants. Interviews were tape-recorded and transcribed for analysis. For each site, one of the researchers summarized responses to the guiding questions across interviewees.

Descriptions of each program's evaluation structures and processes were derived from the on-site interviews. Researchers also reviewed program documents, such as handbooks for university supervisors and cooperating teachers, for additional details and to corroborate information provided by interviewees.

Interviewees reviewed the evaluation program description developed for their particular site. They were asked to check for accuracy and to indicate their permission to quote from the interview. The second researcher on each site team also reviewed the description

for the site. Descriptions were revised to reflect comments from interviewees and researchers. The individual descriptions for each site are provided in this report's companion document, *Systematic Evaluation for Continuous Improvement of Teacher Education: Volume 2: Case Summaries* (Lauer & Dean, 2003).

Researchers at McREL compared responses to the guiding questions across the four sites and identified common themes. This report documents those common themes as well as unique structures or processes that reflect the context of a particular site. All members of the research team and a panel of external reviewers reviewed the cross-case analysis. The report was revised to reflect reviewers' feedback.

RESULTS OF CROSS-CASE ANALYSIS

This section begins with a discussion of study findings, organized around the six questions that guided this study. Included are descriptions of similarities and differences among the programs in the evolution of their evaluation systems, types of data they collect, their use of data, their relationships with P–12 stakeholders, and the culture that supports their collection and use of data. The discussion also highlights themes that emerged in relation to how the sites establish and use structures and processes for systematic evaluation. This discussion is followed by a summary of the findings across the sites to provide an answer to the overall research question.

These four programs have achieved admirable success in collecting and using data for program improvement. Nonetheless, some aspects of these processes continue to challenge them. Based on these challenges, the last part of this Results section suggests areas for further study.

The four teacher education programs reviewed for this study differ in context in a number of ways — including overall size — yet each has established structures and implemented processes that help ensure the effectiveness of its program. Exhibit 1 summarizes these similarities and differences.

Readers will note that three of the four programs are located at private institutions and two of the programs have relatively low numbers of graduates per year. Some readers may wonder to what extent these characteristics influenced the ability and inclination of the programs to conduct systematic evaluation. Galluzzo and Craig (1990) remind us that program evaluation takes place within a social context that is influenced by the political atmosphere, the professional relationship between the program operator and evaluator, the policy environment, and the values held by those who receive the evaluation results. Thus, size and public/private status of the program are important in terms of how they relate to these factors and do not, in and of themselves, determine whether a program will engage in systematic evaluation.

EXHIBIT 1. CHARACTERISTICS OF THE FOUR INSTITUTIONS

CHARACTERISTIC	*ALVERNO*	*EAST CAROLINA*	*FORDHAM*	*SAMFORD*
Private or Public	Private	Public	Private	Private
Total Enrollment, 2001–2002, undergraduate and graduate students	1,950	19,400	14,600	4,400
Make-up of Student Body	Women only	Men & women	Men & women	Men & women
Percentage Diversity	40%	19.7%	25%	12%
# of Graduates, per Year, from Program Receiving Award	20 graduates per year	13 graduates per year	135 graduates per year	60–75 graduates per year
Focus of the Program	Urban	Urban, suburban, and rural	Urban	Urban, suburban, and rural
Level of Teacher Education Students	Undergraduate programs	Undergraduate programs	Graduate programs	Undergraduate programs

Guiding Question #1: How Are Individuals, Groups, and the Components of the Teacher Preparation Programs Evaluated?

The four programs that were the focus of this study use multiple sources of data to gather information about the performance of individuals and groups (i.e., candidates, graduates, faculty members, university supervisors, and cooperating teachers) as well as program components, such as field experiences and content courses. For candidates, the primary sources include performance assessments, portfolios, standardized tests, grade point average, samples of P–12 students' work or achievement, and observation and feedback forms completed by cooperating teachers and supervisors. For graduates, although the primary source of feedback is questionnaires, P–12 student achievement data and informal conversations with graduates also are sources. For data about the performance of faculty members and university supervisors, primary sources include questionnaires and informal feedback from candidates and cooperating teachers. Similarly, each program relies heavily on formal and informal feedback from candidates and university supervisors to make judgments about the program's components, including field work, courses, and the effectiveness of cooperating teachers.

Candidates. Performance assessments are a key source of information about individual candidates' levels of knowledge and skill for all four programs, although the emphasis on their use varies across the sites from about 50 percent at ECU and Fordham to 100 percent at Alverno. Each program also requires candidates to complete portfolios that provide evidence of a range of skills, including candidates' abilities to address the needs of diverse learners and to reflect on their teaching experiences and practice. Portfolios are evaluated at various points in the program (e.g., before admission to student teaching at Alverno) and during the final semester for evidence that the candidate has met program criteria. A unique feature of Alverno's system is an interview conducted during the second semester of education course work to determine candidates' growth in the ability to self-assess.

Three of the four programs require a standardized assessment of candidates' content and/or teaching knowledge. East Carolina University requires its candidates to take Praxis II; Fordham candidates complete the Liberal Arts and Sciences Test and the Assessment of Teaching Skills-Written; and Samford candidates complete a Major Field Assessment Test at the end of their coursework. Alverno candidates demonstrate content knowledge by reaching advanced levels of proficiency in the content-area courses required for completion of their teacher education program.

Another important data source for ECU, Fordham, and Samford is candidate grade point averages (GPAs). All three institutions regularly monitor GPAs. At Samford and ECU, candidates who do not maintain at least the minimum GPA are dropped from the program. At Fordham, candidates who fail a course sometimes have to re-take the course or complete extra work. Alverno candidates do not have GPAs since Alverno does not give grades. Instead, Alverno monitors candidates' progress toward required levels of performance on eight abilities through performance assessments in all courses.

Samples of work from candidates' P–12 students are another data source for all of the programs. These samples are gathered as evidence of candidates' teaching effectiveness during the student teaching experience and occasionally as part of assignments for other courses. For example, Alverno candidates submit samples of student work and an analysis of the lesson during which students produced the work. Faculty use a rubric to review the work sample and provide candidates with feedback about their performance in relation to the abilities that frame the teacher education program. In the past, Alverno faculty also have used these work samples to gauge the effectiveness of their program. They reviewed all of the work samples from several years and looked for patterns in candidates' performance. They found that candidates' analysis tended to focus on their behaviors as teachers (e.g., clearly stating objectives, engaging students' interest) rather than on students' learning. This helped the faculty understand the need to clarify directions and led to changes in the rubric for reviewing candidates' work samples.

Fordham uses candidate work samples primarily to gather evidence of individual candidates' knowledge and skills. Work samples across cohorts of candidates have also been examined for evidence of ways in which high-achieving and low-achieving candidates have met Fordham's internal standards. A set of vignettes was written to capture this evidence.

All of the programs require university supervisors and cooperating teachers to complete observation and feedback forms about candidates' performance in the classroom during field experiences. Cooperating teachers also provide informal feedback about candidates through conversations with university supervisors.

Graduates. For a view of graduates' teaching effectiveness, all of the programs conduct surveys of graduates and principals of the schools in which graduates teach. ECU and Samford also collect achievement data for graduates' P–12 students. In ECU's case, graduates gave permission for the program to look at their students' scores on state-required assessments. Samford has examined the achievement results of its graduates' students on the Stanford Achievement Test in reading and mathematics, QRI reading, and S.T.A.R. reading. Fordham faculty members collect information about the performance of graduates' P–12 students as part of their research program in P–12 schools, but it is not a formal part of Fordham's evaluation system. The programs also collect data through informal conversations with graduates who enroll in seminars for new teachers, through meetings of professional organizations, and as faculty members have contact with graduates in local schools.

Faculty Members and University Supervisors. The primary sources of information about the effectiveness of faculty members and university supervisors are questionnaires completed by candidates. Information about the performance of faculty members are generally gathered by having candidates complete standard questionnaires used university-wide for faculty and course evaluation. These evaluations usually address adequacy of course content, amount and frequency of feedback, respect for diverse talents and ways of learning, opportunities for student-faculty and student-student interaction, and the quality of instruction. Candidates in all four programs complete feedback forms that address the performance of their university supervisor. In general these forms ask candidates to rate how well the supervisors carried out their duties (e.g., clarified the program's expectations, observed an appropriate number of times) and assisted the candidate in improving instructional practices and planning, evaluating, and managing the classroom. Information also is gathered informally through conversations with candidates.

Program Components. All of the programs gather feedback from candidates about courses and field experiences. For example, candidates in ECU's program complete questionnaires after each field experience and at the end of the final internship. They also

meet as a group with the university supervisor several times throughout the semester and provide feedback about cooperating teachers and other aspects of the program. All of the institutions collect data about cooperating teachers informally through conversations with candidates and through university supervisors' interactions with cooperating teachers when supervisors are in the schools in which candidates are placed. Interviewees from all of the programs stressed the importance of this informal data collection.

Other evaluation data about program components is gathered from university supervisors, cooperating teachers, program graduates, and principals of graduates. University supervisors collect data about the field experiences component of the program through the use of specified forms as well as through informal conversations with principals and cooperating teachers. For example, supervisors gather information about the extent to which cooperating teachers fulfill their responsibilities for planning with the candidate, providing feedback to the candidate, and preparing required reports. Representatives from all four programs highly valued these conversations as a source of information about how well the program is functioning. Cooperating teachers also complete specified evaluation forms about the program, providing feedback about how well the university supervisor carries out his or her duties (e.g., maintaining a cooperative relationship with the candidate and the cooperating teacher, preparing required reports, and clarifying expectations of candidates and cooperating teachers). At ECU, cooperating teachers provide additional feedback in focus group sessions conducted twice a year by the clinical coordinators. Samford conducts similar focus groups with P–12 teachers and principals.

All of the programs conduct formal annual surveys of graduates and principals of graduates to determine the effectiveness of various aspects of the program. These surveys address perceptions of graduates' preparedness to carry out different aspects of teaching (e.g., classroom management, content knowledge, needs of diverse learners). For example, the ACT survey used by Samford asks candidates to rate the course content and instruction in their major field, the attitude of faculty toward candidates, the availability of their advisor, and preparation for their future career. Among other questions, Alverno's Graduate Survey asks graduates to assess the extent to which the program prepared them to promote student learning, teach critical thinking, and assess student learning.

Although the source of the surveys differs at each school — for example, Fordham's is custom designed to link to the program's goals and mission, Alverno's is a modification of a tool used by the National Center on Restructuring Education, Schools, and Testing (NCREST)— each of the programs highly values the information provided by the surveys.

Samford augments the information gathered through its survey by conducting focus groups with graduates to identify in greater detail the needs and problems raised in

surveys and to determine the likelihood that proposed solutions will be effective. ECU biannually conducts interviews with graduates.

Other sources of information about program components include feedback from advisory groups (e.g., Samford's Employers Advisory Group, Alverno's Teacher Advisory Council) and feedback from faculty groups (e.g., Samford's Beginning Teacher Team, Alverno's Council for Student Assessment). First-year teachers at Samford and Alverno provide additional feedback informally during their participation in seminars for first-year teachers. Although these sources of data are primarily informal, the programs find that they provide valuable information for making decisions about which changes should be made to the program.

Guiding Question #2: How Do Teacher Preparation Programs Align Evaluation with Program Standards/Goals?

For three of the four programs (Alverno, ECU, and Samford), program goals reflect standards promoted by NCATE, INTASC, the National Board for Professional Teaching Standards, and the institutions' state standards. Fordham aligns its goals with NCATE and state standards. ECU's and Samford's goals are also aligned with other national organizations (i.e., National Council of Teachers of Mathematics, the National Council for Exceptional Children). Samford also incorporates the principles and standards of the Alabama Reading Initiative.

For each of these programs, program goals provide a framework for evaluation. The programs align evaluation data with their goals using candidate performance data and data on program components. For example, at Alverno, candidate performance on reflection logs and portfolio assessments are examined to determine if candidates are acquiring the knowledge, skills, and dispositions outlined by the goals of the program. Similarly, at ECU, portfolio assessments and feedback from candidates, university supervisors, and cooperating teachers on program components provide evidence that the program's goals are being achieved. In addition to looking for evidence in candidate portfolios, Fordham uses data collection forms that are aligned with its program goals to assess candidates' performance in field placements and student teaching. At Samford, faculty members collect data on a set of 14 key performance indicators to determine if program goals are being met. Faculty curriculum committees meet each semester to review the data collected on the key indicators and make decisions about changes based on these data.

Guiding Question #3: How Do Teacher Preparation Programs Develop Systematic Evaluation?

Systematic evaluation is a purposeful activity that includes defined roles and responsibilities for a variety of stakeholders. It involves establishing leadership, determining a purpose, selecting information sources, developing or choosing instruments, establishing a schedule, collecting data, analyzing results, reporting and disseminating results, and implementing follow-up actions (Brainard, 1996). This study examines the different ways in which the approaches the four programs took to systematic evaluation developed. For some of the programs, particular events led to the development or improvement of structures and processes for evaluation. For other programs, the actions of one or two key people were the catalyst for change. For Alverno and Samford, the catalysts were internal to the system. For ECU, the catalysts were primarily external, and for Fordham, there were both internal and external catalysts.

Although the source and nature of each program's catalyst for evolution of the evaluation system were different, the end result was the same — a system for continuous improvement of the teacher preparation program. In all four cases, the system of evaluation that exists today is purposeful, comprehensive, and institutionalized. Alverno provides an example of how a system might change when the catalyst is system-wide and internal. For Alverno, the initial catalyst was conversations about the curriculum initiated by the president of the college. These conversations led to the creation of an ability-based curriculum in the 1970s, which caused faculty to think differently about teaching, learning, and assessment. This was the beginning of Alverno's emphasis on performance assessment and the Education Department's efforts to define a conceptual framework around the abilities to guide instruction and assessment in the teacher preparation program. Since the 1970s, Alverno faculty members have been refining their understanding of the ability-based curriculum — how to teach it, how to assess students' progress relative to it, and how to continuously monitor it for improvement. Their approach to system evolution has been deliberate and purposeful.

Samford provides another example of a system-wide, internal catalyst for change. Samford began by developing performance indicators to monitor its program. Performance on some of the indicators served as the impetus for the collection and use of data to improve the program. At about the same time the indicators were developed (1994), Samford adopted problem-based learning as its approach to the curriculum. This spurred faculty to have conversations about the redesign of the curriculum and to develop a model to guide changes. Each step of the model includes data collection; these data are used to draw conclusions about the effectiveness of the program. Samford's change process is inclusive — it uses input from faculty, candidates, graduates, P–12 practitioners, and employers.

In both the Alverno and the Samford cases, a system-wide curriculum focus made it necessary — and easier — for faculty within and across departments to work together. In the process, faculty developed new ways of interacting and models for change that included an emphasis on data collection and use.

East Carolina University is an example of how a system developed when the catalysts for change were positive, external factors, including new faculty who brought different ideas, expectations, and ways of working to the program. External factors resulted in leadership and structures that made it easier for the group to work together toward a common goal. When a new middle school coordinator was hired at ECU in 1994, her experiences in other programs and her willingness to listen led to changes in the program, including the addition of performance-based assessments in all courses and a requirement that candidates compile a portfolio to demonstrate their knowledge and skills. When another new faculty member with a similar philosophy was hired, more changes occurred. At about the same time, the Department of Mathematics received a grant from the National Science Foundation to redesign the middle grades mathematics courses and the state provided the university with funds to establish partnerships with P–12 schools. The latter led to the establishment of the Clinical Schools Network (CSN), a partnership between the university and P–12 school districts. The CSN strengthened connections between teacher education and the P–12 schools and led to a number of improvements in the program, including the development of specific forms for data collection about various aspects of the program (e.g., roles and responsibilities of supervisor, clinical teacher, and candidate).

The impetus for revisions to the program at Fordham over the years has been a combination of external influences, such as NCATE and state regulations, and internal influences, primarily faculty observations of the need for change. Specifically, the loss of NCATE accreditation in the early 1990s was the motivation to regularize processes involved in tracking academic data. The reinstatement of NCATE accreditation in 1996 spurred the development of new structures for data collection, such as a graduate survey, and the systematization of processes for data collection, such as the use of assessment rubrics. Over the years, collection of data on candidates has become more formal, more systematic, and less anecdotal. For example, new forms for supervisors and a special form for observing a lesson have been developed. The use of rubrics for candidate performance is an area that the program has targeted for additional evaluation development. Also, the school of education currently is developing a process for evaluating employer satisfaction, and a university committee on quality teaching is looking at alternate ways of measuring the quality of faculty teaching performance, perhaps using rubrics.

The differences among the four cases demonstrate that there are a number of ways to develop systematic evaluation for program improvement. Although university-wide

programs provide a common focus and support for data collection efforts, individuals can also build evaluation systems by making improvements to the parts of the system they directly affect and by recognizing and leveraging other opportunities that support their efforts. In each of the cases, the evaluation system moved forward when the various stakeholders had opportunities to work together to collect and use data for making decisions about program improvement.

Institutionalization of the Evaluation System. From the above descriptions, it is clear that development of evaluation by the four programs occurred in stages or phases. Catalysts spurred fundamental changes in structures and processes, followed by refinement, until another key event or person catalyzed the system and forced another fundamental change. All of the programs can cite evidence of institutionalization, such as formal descriptions about evaluation in documents, routine timelines for data collection activities, funding and time allocations for evaluation activities, shared faculty understandings of candidate assessments and performance, significant input from graduates, and other universities seeking information about the program.

In addition to the above signs of institutionalization, program representatives mentioned several others. For example, Alverno representatives cited identification of the elements of the program that are generalizable or transferable to another institution as evidence that their evaluation program was institutionalized. Fordham representatives mentioned the increased emphasis on the assessment of faculty instruction and faculty-developed teaching assessments in portfolios for reappointment and tenure as a mark of institutionalization. Although each of the programs can cite indicators of institutionalization of their evaluation system, none of the programs is content to remain unchanged; rather, each is seeking new and improved ways to conduct evaluation for program improvement.

Guiding Question #4: How Do P–16 Stakeholders Influence Evaluation of the Teacher Preparation Programs?

P–16 stakeholders influence evaluation of the four nationally recognized teacher preparation programs in several ways, including providing formal and informal feedback about program components and program graduates. They also influence evaluation by serving as external assessors, participating in research and other projects with program faculty, and serving on advisory and curriculum committees.

Although each of these ways of influencing evaluation is important, the primary way that P–16 stakeholders influence program evaluation at all four institutions is by providing feedback about program components (e.g., content of courses and field experiences, training for cooperating teachers) and program graduates (e.g., candidates' content knowledge, ability to manage a classroom, ability to teach specific content such as

writing or reading, ability to teach diverse students). Feedback from P–12 stakeholders is gathered in formal and informal ways through the institutions' partnerships with P–12 schools and districts. For example, at ECU, members of the Clinical Schools Network, a university-school partnership of 15 school districts in eastern North Carolina, provide feedback about the program during the group's monthly meetings and through frequent interactions with the clinical coordinators and university supervisors on site, by phone, or by e-mail. Clinical teachers also complete an evaluation of the teacher candidate and his or her university supervisor. Teachers also are queried about any problems that are occurring, how things are going, and what the university can do to make the experience better for them as clinical teachers or for the teacher education candidates.

P–16 stakeholders also provide feedback through informal conversations that take place during meetings of higher education faculty (e.g., university supervisors, program faculty, curriculum committees) and between program faculty and principals and cooperating teachers in schools where program candidates are placed for field experiences. In addition, as mentioned previously, cooperating teachers, principals, university supervisors, candidates, and graduates provide feedback by completing various observation forms or questionnaires.

In addition to formal partnerships with P–12 schools, all four programs have strong relationships with principals and cooperating teachers in schools where their candidates are placed for field experiences. These relationships are possible because faculty spend a great deal of time in the P–12 schools — supervising student teachers, working with candidates in pre-student teaching field placements and practica, conducting research, providing workshops for teachers and administrators, and working on projects with school staff. Principals and cooperating teachers confirmed the strong presence of the programs in the schools. Program directors and field supervisors emphasized that the informal conversations they have with principals and cooperating teachers are an important way to collect data about the effectiveness of the teacher education program. Principals interviewed commented on their close relationships with program directors and field supervisors. They described the openness of program directors and faculty to feedback and the willingness to make improvements.

P–16 stakeholders influence evaluation in other ways. For example, Alverno involves principals and teachers as external assessors who review candidates' portfolios before they enter student teaching. Similarly, Fordham's Holmes partnership committee involves faculty and P–12 teachers and administrators in conducting research that will lead to improved teaching and teacher education. At ECU, the teacher-in-residence program provides another way for P–12 stakeholders to influence evaluation. The teacher-in-residence in the Department of Mathematics helps collect data for the program and serves as a resource for mathematics education faculty by providing feedback about course content and teaching methods and for candidates by commenting on their lesson

plans. P–12 practitioners also influence course development at ECU by participating on advisory and curriculum committees or special task forces and working directly with faculty to develop courses.

Guiding Question #5: How Do External Influences Affect Evaluation of the Teacher Preparation Programs?

For all four programs, the primary external influence is state policy. State policies on teacher preparation are sources of program revision and provide impetus for evaluation activities, such as conducting surveys of graduates or establishing advisory committees that include P–12 representation. The programs are proactive in ensuring that candidate preparation is aligned with both current and anticipated changes in state regulations. For example, Samford aligned its assessments of candidates' field experiences with the criteria on the Professional Education Personnel Evaluation, the first-year teaching evaluation required by the state of Alabama. Similarly, ECU now requires candidates to complete a portfolio that prepares them to meet the state's portfolio requirement for certification.

All of the programs mentioned that they are affected by the national emphasis on accountability. Fordham representatives specifically mentioned Title II of the Higher Education Act's requirement to report pass rates of graduates on state licensure tests. (Wisconsin did not have a state licensure test requirement at the time this study was conducted.) Of the four programs, Alverno emphasized most strongly that the evaluation of its program is influenced by other higher education institutions. Alverno has participated in a number of partnerships that have focused on assessment of learning, such as the Consortium for the Improvement of Teaching, Learning and Assessment (1989–1992), and the Teaching for Tomorrow Project (1998). Through these partnerships, Alverno has benchmarked itself against the approaches that other institutions take to higher education in general and to teacher education in particular. As mentioned previously, all of the programs align with various state and national teaching standards, accreditation standards, and content-area standards.

Guiding Question #6: What Are the Characteristics of a Culture that Supports Data Collection and Use for the Evaluation of the Teacher Preparation Programs?

These four programs are distinguished by their commitment to collecting and using data for program improvement. As described previously, this focus on data did not develop overnight. There were changes in attitudes, relationships, skills, and practices — the culture — that had to occur before data collection and use became part of the fabric of the program. The characteristics of a culture that supports data collection and use for program improvement are an attitude that data are essential, training in using data, time to

analyze and interpret data, practices that encourage involvement and build commitment to evaluation, collaboration within and across departments, and a commitment to helping candidates succeed (Ayers, 1981; Galluzzo & Craig (1990); Holcomb, 1999).

An Attitude that Data Are Essential. One of the most important characteristics of a culture that supports data collection and use is an attitude that data are essential for gauging progress and making decisions that affect the program. Program faculty at each of these sites view data collection and use as part of a continuous improvement process. Further, data aren't collected simply to satisfy an external need, such as NCATE accreditation, but primarily because there is a commitment to having the best teacher preparation program possible. There is an expectation that everyone connected with the program is striving for excellence. In Alverno's case, the commitment is "bone deep" — faculty and staff are in constant pursuit of a deeper understanding of how they can help candidates acquire and apply the knowledge embedded in the ability-based curriculum. Similarly, at Samford, using data to make decisions is very much a part of the university's culture — it's how everyone works together.

In order to hold positive views about data collection and use, faculty must be shown that it is "safe" to examine data on the results of their work. Each of the programs has developed a culture of trust that the data will be used to help people rather than to "get" them. These programs built trust by providing faculty with opportunities to give input and feedback on evaluation activities and program changes. They also involved faculty in collaborative development of assessment instruments that are used across the program. This approach helped overcome faculty members' perceptions that they were losing their academic freedom and independence to make decisions about measuring candidates' performance. Even though it can take considerable time to reach consensus when using a collaborative process, the end result — a sense of shared ownership — is worth it.

Training in Using Data for Evaluation. Training also can increase trust and faculty members' confidence in using data for evaluation. Representatives from three of the programs mentioned receiving specific data-related training. For example, Samford provided faculty with training in benchmarking (i.e., comparing a program to another that is considered of high quality) and in designing their own teaching portfolios. Fordham faculty received training on rubrics. At Alverno, all faculty members have opportunities to deepen their understanding of the abilities and how well candidates are learning them during the Tuesday afternoon department meetings since how the abilities are taught and assessed is the usual topic for these meetings. The same is true, to a varying degree, for the Friday afternoon faculty meetings and the college-wide faculty institutes, which are held three times a year.

Time to Discuss and Analyze Data. Another characteristic of the culture at these sites is designated time to discuss and analyze data. The faculty at each of the institutions

studied meets regularly — weekly, monthly, and/or at annual retreats. For example, at Fordham, survey and other data in the school's annual report are discussed at a year-end annual retreat. Instructors meet with program coordinators at the beginning and end of courses to give feedback. Program coordinators meet as a group every two weeks. At the meetings, coordinators discuss needed program revision and then give proposals for the revision to program faculty. Coordinators meet with field supervisors monthly in person or by phone, at least monthly. At ECU, the middle school program coordinators meet on an almost daily basis to discuss the program and hold retreats twice a year to examine data in more detail.

Incentives that Encourage Involvement and Build Commitment to Evaluation. Also characteristic of the cultures of each of the sites is attention to incentives to encourage involvement and to build commitment to evaluation activities. The incentives vary somewhat, but each is intended to provide recognition or reward for collecting and/or using data to make program improvements. A common incentive among the programs is aligning criteria for promotion and tenure with participation. For example, at Fordham and ECU, faculty members are required to conduct and publish research as well as demonstrate excellence in teaching and provide service to the education community. Faculty members are encouraged to conduct their research on the teacher education program, thereby generating information for program improvement. At Samford, faculty can demonstrate scholarly activity by developing teaching portfolios that emphasize the improvement of instruction through data collection. For Alverno faculty, another incentive is the need to validate the ability-based curriculum. Opportunities to travel to conferences and interact with colleagues from other institutions and national organizations and the satisfaction of seeing the program improve as a result of evaluation activities also serve as incentives for program faculty at the four sites.

Collaboration Within and Across Departments. High degrees of collaboration within and across departments is an important characteristic of the culture at each of these sites. Collaboration among teacher education faculty occurs through formal and informal meetings and course development. At all four institutions, faculty members serve as mentors. At Fordham and Alverno, they also serve as peer reviewers. At Alverno, collaboration also includes co-teaching of some classes.

Each of the programs fosters collaboration across departments — in particular with departments in the liberal arts and sciences. For three of the programs (Alverno, Fordham, and Samford), this was accomplished by establishing a committee for the specified purpose of increasing the level of conversation about important issues of course content and delivery. For example, Fordham established a teacher education executive committee (TEEC) to facilitate communication and strengthen connections between arts and sciences and teacher education. At ECU, mathematics education faculty members are

not part of the School of Education. They meet formally and informally, however, with members of the middle grades program, which is in the School of Education. Teacher education and arts and sciences faculty at these four institutions also collaborate through co-design or development of courses. For example, at Fordham, arts and sciences faculty members are involved in the design of methods courses for teacher education.

Commitment to Helping Candidates Succeed. Faculty members at these institutions share a commitment to helping candidates succeed, and a central component of success is being able to have a positive impact on P–12 student learning. This commitment to candidate success is one reason faculty members are willing to participate in evaluation activities. They want to know whether the program is effective, and if it isn't, they want to make it better. For example, at Samford, program evaluation is driven by the program's clients — graduates and their employers. Samford's goal is that every candidate who seeks a teaching position is hired after graduation and that employers judge their performance during their first year of teaching as successful. For this reason, first-year graduates and the principals who hire Samford graduates are primary sources of input about Samford's program.

Representatives from Alverno and ECU mentioned that they use the hiring process to select faculty who have a commitment to helping candidates succeed. For example, at ECU, faculty in the award-winning program and the dean of the School of Education check during the hiring process that candidates agree with the agenda the dean and other program leaders are pushing — partnerships with P–12 schools, for example. Alverno makes clear to job applicants that to be hired, they must be dedicated to teaching rather than to their individual research, willing to work across disciplines, and willing to engage in collaborative work. As mentioned previously, these characteristics contribute to faculty's willingness and ability to contribute to program evaluation.

Summary of Structures and Processes — Answering the Overall Question

This section summarizes information from the previous section to answer the overall research question about the structures and processes that the four nationally recognized teacher preparation programs use to determine and improve the effectiveness of their programs. For this study, "a structure" refers to an element of the evaluation system that helps the program collect, analyze, or use data to improve the program. For example, a partnership between the teacher education program and P–12 schools is a structure. One of its functions is to serve as a mechanism for collecting feedback about the effectiveness of the program. Structures of systematic evaluation of the programs in this study are listed in Exhibit 2.

EXHIBIT 2. STRUCTURES OF SYSTEMATIC EVALUATION USED BY THE FOUR NATIONALLY RECOGNIZED TEACHER PREPARATION PROGRAMS

- Data-collection strategies
- Meetings and trainings
- Partnerships (including participation in consortia or projects with groups outside the teacher preparation program)
- Offices, committees, or advisory groups
- External guidelines and standards
- Program goals, frameworks, or indicators

One definition of a *process* is a systematic series of actions directed toward an end. In this study, the word is broadly interpreted to include a series of actions such as hiring and less well-defined actions such as communication. Processes used in systematic evaluation by the four programs are listed in Exhibit 3.

EXHIBIT 3. PROCESSES USED IN SYSTEMATIC EVALUATION BY THE FOUR NATIONALLY RECOGNIZED TEACHER PREPARATION PROGRAMS

- Use of data
- Collaboration
- Hiring
- Communication
- Faculty evaluation
- Continuous improvement/change process
- Shared leadership

Structures. This section provides summary statements and examples regarding the structures of systematic evaluation used by the four teacher preparation programs examined for this study.

The four programs studied use a variety of data collection strategies, including informal ones. As noted in previous sections, common data collection strategies include graduate and principal surveys, performance assessments, portfolios, and focus groups. Informal collection of feedback also is characteristic of each of the programs. All of the programs highly value the informal conversations that program faculty have with P–12 teachers and principals and consider them a significant

avenue for gathering information. Each of the programs worked hard to establish good relationships with P–12 partners so that such conversations could be a meaningful source of data. The programs have developed formal feedback loops, such as regular meetings among various program faculty members, to support these informal conversations.

Each of the programs has offices that carry out various functions that support its evaluation program. These include the Assessment Center and the Educational Research and Evaluation office at Alverno; the Office of Teacher Education and the Institutional Planning and Research Office at ECU; the Office of the Dean of the Graduate School of Education at Fordham; and the Office of Assessment at Samford. These offices assist in the development of data collection tools and in storage or analysis of data. Formal committees and advisory groups, such as the Council for Student Assessment and the Teacher Advisory Committee at Alverno, the Teacher Education Executive Committee at Fordham, faculty curriculum teams at Samford, and the Council for Teacher Education at ECU, provide input and feedback on program components and help design and implement program improvements.

Formal meeting times are set aside to support the use of data for program improvement. Teacher education faculty attend meetings designed to provide them with opportunities to give input about the collection of data, to conduct analyses of data, and to discuss ways to use the results of the analysis to guide program improvement. Having specified times and a program improvement focus for these meetings are key to making these meetings a meaningful part of the evaluation system. Three of the four programs provided training for faculty members to help them understand the need to collect data on various aspects of the program and to develop skills in analyzing and using data. In addition to accomplishing these goals, training also built a level of trust among faculty members that allowed evaluation activities to become an important part of the life of the program.

Partnerships with K–12 schools are used as a source of data for determining the effectiveness of the teacher education program. Hass and Stoffels (1998) found that teacher education programs undergoing reform, and implementing evaluation activities as part of that reform, had to develop internal and external relationships to facilitate their efforts. This was certainly true for the four programs in this study. All four programs have established partnerships that provide access to feedback from P–12 school stakeholders. Principals and cooperating teachers commented on their close relationships with program directors and field supervisors, noting the openness of program faculty to feedback and their willingness to make improvements. Partnerships with P–12 schools aren't limited to professional development schools or networks that provide a way for P–12 practitioners to provide input to the program. They also include partnerships that focus on co-development of courses for teacher

candidates and assessment projects. Such partnerships provide program faculty with firsthand knowledge of P–12 issues and the types of skills that graduates will need to function effectively in the P–12 system.

Standards, goals, and principles are used to guide the evaluation of the program. Many sets of standards and principles guide the programs. As described previously, standards from NCATE, INTASC, the National Board for Professional Teaching, and the states influence the design and evaluation of each of the programs. In addition to standards, each program is guided by a clearly articulated set of goals and a conceptual framework that outlines the elements and principles of the program. The programs use a variety of representations, including conceptual maps and matrices, to increase understanding of the goals of the program and the way the pieces of the program fit together. Campus-wide initiatives also play a role in the design and evaluation of two of the programs. At Alverno, the teacher education program is designed to ensure that candidates acquire abilities that have been defined as the targets for learning across the college. At Samford, the teacher education program, like all other programs on campus, is built on the principles of Total Quality Management and Problem-based Learning.

A strong, coherent curriculum that is aligned with program goals drives data collection and determinations of program success. The coherence of the curriculum is evident in the common language and understanding among faculty and in documents that describe the program. For example, in interviews for the study, teacher education faculty and program directors used the same language to describe how their curriculum connects to program elements and goals. Coherence also was evident in written materials, such as Samford's course matrices that illustrate the connections between the curriculum and the four program threads (problem-based learning, diverse populations, clinical experiences, and technology) as well as NCATE and National Board Standards. Similarly, Alverno's visual representation of how the five abilities (i.e., conceptualization, diagnosis, coordination, communication, and integrative interaction) link to a set of key concepts (i.e., developmental needs, diversity, professionalism, and media/technology) and essential processes in which teachers engage (i.e., planning, implementing, assessing) clearly illustrates and helps maintain the coherence of the curriculum.

Processes. This section provides summary statements and examples regarding the processes of systematic evaluation used by the four teacher preparation programs examined for this study.

Most data is used in a formative way — to make changes to the program on a regular basis. As mentioned previously, to receive the award, programs had to provide three types of evidence of program effectiveness: formative, summative, and confirming. The programs reported that they used most of their data in a formative

way. In fact, many of the data sources identified as summative sources of evidence (e.g., end-of-program portfolios, Praxis II scores) or confirming evidence (e.g., P–12 student achievement data) also were used formatively. This means that these sources were used to make decisions about changes to the program and were not used strictly to measure candidates' knowledge, skills, and dispositions at the end of courses, field experiences, or the program or to determine graduates' effectiveness in P–12 classrooms.

Collaboration and communication, especially from and to leaders, are keys to the evaluation system. Faculty members from all four programs emphasized the important role that communication and collaboration play in their evaluation systems, although the levels of collaboration varied among the programs. Faculty members in each program indicated that they collaborate on the use of data for redesigning programs and curriculum, and deans and program directors described this collaboration as essential to sustaining an evaluation system.

Effective collaboration is not possible without good communication. Each of the programs fosters and supports good communication through the structures described in the previous section. In addition, program and institutional leaders are accessible, seek input from a variety of stakeholders, welcome feedback, and value partnerships. They set expectations for excellence and for questioning the status quo.

Hiring and faculty performance evaluation practices are used in ways that support the evaluation system. The four programs in this study have learned the importance of hiring faculty who value norms of collaboration and evidence of effectiveness. The programs highlight these qualities when advertising for faculty positions and when conducting interviews. In addition, their faculty performance evaluation criteria are linked to these qualities. For example, Alverno has a faculty performance evaluation criterion that focuses on working responsibly in the college community, which encourages participation in evaluation and improvement of the program. The other three programs are at various stages of development of similar faculty performance evaluation systems.

A continuous improvement approach is used. Many institutions view continuous improvement as an activity in itself, but as Hass and Stoffels (1998) found, teacher preparation programs implementing evaluation activities as part of reform have to view continuous improvement as the goal of evaluation. These four programs understand this point. They view the work of program improvement as an ongoing process and not something that is the focus of attention only for defined periods of time, such as an NCATE review. They find ways to help faculty members understand and use a continuous improvement approach. For example, Samford developed a model and principles to guide its change process — the Essential Changes model and

the Quality Principles of Change. Together, these encourage teamwork and a problem-solving approach to improvement.

Leaders share responsibility for improvement. In varying degrees, program leaders in each of the four programs share leadership by facilitating the involvement of faculty in the collection and use of data. Samford typifies the approach used at three of the institutions; faculty describe leadership as facilitative, meaning that program leaders bring data and research findings to the faculty and other stakeholders and then facilitate the use of this information for program improvement. At Alverno, leadership for improvement is shared extensively; *all* faculty (the chair of the Education Division, field supervisors, professors for content courses, and methods instructors) are involved in gathering and analyzing data.

The structures and processes used by the four programs in this study demonstrate what it means for systematic evaluation to be a purposeful activity. These programs are dedicated to preparing teachers who are effective in helping all students learn. They accomplish this task by having clear standards and principles that guide all aspects of their program. But these programs know that having standards and principles isn't enough. They know that they must continually monitor progress toward meeting those standards by gathering, analyzing, and interpreting a variety of data and using the results to improve their program. And they know that strong relationships among their own faculty and between program faculty and their K–12 partners are needed to effectively carry out these actions. From these programs, it is clear that systematic evaluation thrives where communication is open and frequent, hiring practices support the goals of the program, and there is shared leadership. These programs demonstrate that systematic evaluation is "doable" and that it makes a difference. But it requires willingness to work collaboratively and commitment to use the results of the evaluation for continuous program improvement.

Limitations of the Study

The applicability of the findings reported in this study may be limited for a number of reasons. For example, the sample for the study was restricted to the four winners of the National Award for Effective Teacher Preparation, which were selected from a relatively small applicant pool. The pool may have been small because it was the first year of the award program and many institutions may not have been aware of the program, did not have time to complete the application, or did not think there were sufficient incentives for applying. (No money was associated with winning the award.)

Also limiting the study is the state of evaluation of teacher preparation. Galluzzo and Craig (1990) note that evaluation of teacher preparation is still in its infancy. Although the programs in this study have made significant progress in developing their systems of evaluation, they acknowledge that there is more to be done. This is particularly true with

respect to collecting confirming evidence — evidence that graduates can help all students achieve high standards. Some argue that limiting confirming data to interviews or survey responses makes it difficult to draw valid conclusions about program effectiveness. Others argue that confirming evidence has little meaning since the context in which graduates teach can greatly influence the extent to which a graduate uses what was learned in the program.

There may be some concern about the applicability of the study because there is disagreement about what it means for a teacher education program to be effective or "proficient." As mentioned previously, although there are standards for teacher preparation to guide the field, there is not universal agreement about how to use those standards for program evaluation. In addition, there is not agreement about the outcomes that should be measured and which of the outcomes should be most important. The award program required applicants to provide formative, summative, and confirming evidence of their effectiveness, knowing that confirming evidence might not be as substantial as the other two categories. This was done in part to get a sense of how programs were addressing confirming evidence and to encourage further conversation about the issues surrounding its collection and use. The issue of confirming evidence is discussed further under Areas for Further Study.

Although a variety of stakeholders were interviewed at each site, the number of interviewees in some categories at some sites may have limited the perspectives provided. For example, researchers had access to few program candidates and graduates. (In some cases, however, cooperating teachers interviewed also were graduates of the program.) Findings also may be limited because two of the programs have a small number of graduates each year. Some may argue that this makes it easier for the programs to gather and analyze data, and to use the results of their evaluation to make program changes. The field would benefit from studies of effective institutions that produce many more graduates each year. Along a similar vein, three of the four programs in this study are in private institutions. Further study is needed of additional public and private teacher education programs to determine how or if public/private status affects the processes, approaches, and level of evaluation.

Despite these limitations, the programs in this study use practices consistent with those recommended by Ayers (1981) and Vaughn (1981), which emphasize a comprehensive approach to evaluation, examining outputs, involving all stakeholders, and using the results for improvement. They also reflect the view of evaluation as a social activity, as outlined by Galluzzo and Craig (1990), and have most, if not all, of the features of the programs of excellence described by Darling-Hammond (2000) and Scannell (1999), particularly a clear framework of principles and goals, a coherent curriculum, and an emphasis on performance assessment. Similarly, Galluzzo and Craig note that "without a collection of purposeful processes on which the faculty members have reached

consensus, there can be no program evaluation that has meaningful implications for program improvement, understanding, and knowledge production." (p. 610). Faculty members of each of the programs in the study have come to agreement on what their candidates should know and be able to do and program evaluations are tied to these agreed-upon standards. As suggested by Thomas and Loadman (2001) these programs also use both qualitative and quantitative methods to obtain a more complete perspective on the various aspects of the program. For example, although the collection of data about graduates' effects in classrooms is still primarily through surveys of graduates and principals, each program has made inroads to other methods of determining how well their graduates help students learn (e.g., student work samples, student performance on state assessments).

This study suggests that these programs have made significant progress in developing structures and processes to monitor and improve the effectiveness of their programs. It also suggests that there are many remaining questions about systematic evaluation of teacher preparation. Several such questions are proposed in the next section.

AREAS FOR FURTHER STUDY

The programs included in this study provide examples of how others can approach the task of systematic evaluation of teacher preparation and reap benefits for graduates of their program and the students they teach. Although there is much to be learned from these programs, there are questions and challenges that remain. These questions and challenges are addressed in this section.

Determining Graduates' Teaching Effectiveness

In this study, the term "confirming data" is used to refer to evidence that is gathered to demonstrate that program graduates are effective in helping all students achieve high standards of learning. Although the programs studied seemed to find it relatively easy to collect data about candidates' performance as they progress through the program (formative data) and at the end of the program (summative data), the programs continue to face a number of challenges in collecting confirming data. These challenges range from difficulty in locating graduates to a lack of consistency in the types of data and test scores reported by P–12 schools. They also reflect the fact that the structures and processes for collecting confirming data are influenced by such contextual factors as whether the state assesses P–12 students and makes the data readily available to schools, the level of sophistication of schools and districts in using state or district assessment data, opinions about what counts as confirming data, and attitudes about whose responsibility it is to collect such data. Possible questions to investigate include the following:

- What forms of data are best for determining graduates' effectiveness in P–12 classrooms?
- What are effective ways to synthesize various types of data about graduates' effectiveness in P–12 classrooms?
- How should the responsibility for data collection and analysis of graduates' effectiveness in P–12 classrooms be shared among teacher education programs, P–12 schools, and the state?

Managing Data

The programs examined for this study vary in their capacity to manage data. Three have staff members whose designated role is to enter and/or analyze data. All acknowledged the need for more assistance, particularly electronic ways to gather and store data about candidate performance. Clearly, without easy access to data, programs are hampered in their efforts to analyze, synthesize, and apply evaluation results. Investigating the following questions will help programs think more carefully about the capacity needed to manage data for program improvement:

- What are effective ways to use technology to collect and analyze data about candidate performance?
- How are faculty members' attitudes toward collecting and analyzing data affected by using technology to collect and use data for program improvement?

Determining Proficiency

In his 2002 report on teacher quality to Congress (U. S. Department of Education, Office of Postsecondary Education, 2002), Secretary of Education Rod Paige emphasizes that teachers' content knowledge is key to their ability to positively affect students' learning. It follows that examining candidates' content knowledge is an important part of the evaluation of a teacher preparation program. Part of evaluating candidates' content knowledge is establishing levels of proficiency. Programs in this study developed their definition of proficient performance by looking at research on best practices, examining data on the performance of prior candidates, and constructing rubrics to describe the relationship between different levels of proficiency and aspects of candidate performance. Their answers to the question of what proficiency means also reflect the standards that guide their programs. The following questions need further study:

- What are some valid and reliable ways to determine candidates' content knowledge?
- What are effective ways to set appropriate levels of proficiency for content and pedagogical knowledge?

If we are to meet the challenge of having a nation of highly qualified teachers, we must ensure that teacher preparation programs are of the highest quality. The four programs examined for this study have committed to meeting this challenge by developing the necessary skills and attitudes to examine their programs from multiple perspectives and to act on the data they collect. By describing the structures and processes that characterize these programs' evaluation systems, this report provides guidance that can make meeting the challenge less daunting.

REFERENCES

American Council on Education. (1999). *To touch the future: Transforming the way teachers are taught: An action agenda for college and university presidents*. Washington, DC: Author. Retrieved from www.acenet.edu/resources/presnet/task-force.cfm

Ayers, J. (1981). Design characteristics for meaningful teacher education evaluation. In Hord, S. & Adams, R. (Eds.), *Teacher education program evaluation 1981: Theory and practice*. Austin, TX: University of Texas at Austin, Research and Development Center for Teacher Education. (ERIC Document Reproduction Service No. ED204 329)

Brainard, E. (1996). *A hands-on guide to school program evaluation*. Bloomington, IN: Phi Delta Kappa Educational Foundation.

Cochran-Smith, M. (2001). Constructing outcomes in teacher education: Policy, practice, and pitfalls. *Education Policy Archives, 9*(11). Retrieved from http://epaa.asu.edu/epaa/v9n11.html

Collias, K. A. (2000). *A report on STEP and its pilot program in Georgia.* Washington, DC: Council for Basic Education.

Darling-Hammond, L. (1999). Educating teachers for the next century: Rethinking practice and policy. In G. A. Griffin (Ed.), *The education of teachers* (pp. 221–256). Chicago: University of Chicago Press.

Darling-Hammond, L. (Ed.). (2000). *Studies of excellence in teacher education.* Washington, DC: American Association of Colleges for Teacher Education.

Diez, M. E. (Ed.). (1998). *Changing the practice of teacher education: Standards and assessment as a lever for change*. Washington DC: American Association of Colleges for Teacher Education.

Galluzzo, G. R., & Craig, J. R. (1990). Evaluation of preservice teacher education programs. In W. R. Houston, M. Haberman, & J. Sikula (Eds.), *Handbook of research on teacher education*. New York: Macmillan.

Hall, G. (1981). What is the future of teacher education program evaluation? In S. Hord & R. Adams (Eds.), *Teacher education program evaluation 1981: Theory and practice*. Austin, TX: University of Texas at Austin, Research and Development Center for Teacher Education. (ERIC Document Reproduction Service No. ED204 329)

Harris, L. B., Salzman, S., Frantz, A., Newsome, J., & Martin, M. (2000, February). *Using accountability measures in the preparation of preservice teachers to make a difference in the learning of all students*. Paper presented at the annual meeting of the American Association of Colleges and Universities, Chicago. (ERIC Document Reproduction Service No. ED 440 926)

Hass, J. M., & Stoffels, J. A. (1998). Supporting change in teacher education institutions. In M. E. Diez (Ed.), *Changing the practice of teacher education: Standards and assessment as a lever for change* (pp. 145–155). Washington, DC: American Association of Colleges for Teacher Education.

Henn-Reinke, K., & Kies, K. M. (1998). Supporting change in teacher education institutions. In M. E. Diez (Ed.), *Changing the practice of teacher education: Standards and assessment as a lever for change* (pp. 157–167). Washington, DC: American Association of Colleges for Teacher Education.

Holcomb, E. (1999). *Getting excited about data: How to combine people, passion, and proof*. Thousand Oaks, CA: Corwin Press.

Holmes Group. (1986).*Tomorrow's teachers — A report of the Holmes Group*. East Lansing, MI: Author.

Howey, K. R., & Zimpher, N. L. (1999). Pervasive problems and issues in teacher education. In G. A. Griffin (Ed.), *The education of teachers* (pp. 279–305). Chicago: University of Chicago Press.

Jordan, H., Mendro, R., & Weerasinghe, D. (1997). *Teacher effects on longitudinal student achievement*. Paper presented at the Sixth Annual National Evaluation Institute, Indianapolis, Indiana.

Lauer, P. A., & Dean, C. B. (2003). *Systematic evaluation for continuous improvement of teacher education: Volume 2: Case summaries*. Aurora, CO: Mid-continent Research for Education and Learning.

Murray, F. B. (2000). The role of accreditation reform in teacher education. *Educational Policy, 14*(1), 40–59.

National Commission on Teaching and America's Future. (1996). *What matters most: Teaching for America's future*. Washington, DC: U.S. Government Printing office.

National Council for Accreditation of Teacher Education. (2002). *Professional standards for the accreditation of schools, colleges, and departments of education: 2002 edition*. Washington, DC: Author. Retrieved from http://www.ncate.org/2000/unit_stnds_2002.pdf

Peterson, K. (1989, summer). Teacher education program evaluation through an institutional board review of graduates. *Journal of Research and Development in Education*, *22*(4).

Sanders, W., & Rivers, J. (1996). *Cumulative and residual effects of teachers on future student academic achievement*. Knoxville: University of Tennessee Value-Added Research and Assessment Center.

Scannell, D. P. (1999). *Models of teacher education*. Washington, DC: American Council on Education.

Tom, A. (1997). *Redesigning teacher education*. Albany, NY: State University of New York Press.

Thomas, A., & Loadman, W. (2001). Evaluating teacher education programs using a national survey. *Journal of Educational Research*, *94*(4), 195–206.

U.S. Department of Education. (2000a). *Eliminating the barriers to improving teaching*. Washington, DC: Author.

U.S. Department of Education. (2000b). *The National Awards Program for Effective Teacher Preparation: Application*. Washington, DC: Author. Retrieved from http://www.ed.gov/inits/teachers/teacherprep

U. S. Department of Education, Office of Postsecondary Education. (2002). *Meeting the highly qualified teachers challenge: The Secretary's annual report on teacher quality*. Washington, DC: Author.

Vaughn, J. (1981). So what is everybody telling us about evaluating teacher education? In S. Hord & R. Adams (Eds.), *Teacher education program evaluation 1981: Theory and practice*. Austin, TX: University of Texas at Austin, Research and Development Center for Teacher Education. (ERIC Document Reproduction Service No. ED204 329)

Wilson, S. M., Floden, R. E., & Ferrini-Mundy, J. (2001, February). *Teacher preparation research: Current knowledge, gaps and recommendations* (Document R-01-3). Seattle, WA: Center for the Study of Teaching and Policy. Retrieved from http://depts.washington.edu/ctpmail/Study14.html

Worthen, B., & Sanders, J. (1987). *Educational evaluation*. White Plains, NY: Longman.

Yinger, R. J. (1999). The role of standards in teaching and teacher education. In G. A. Griffin (Ed.), *The education of teachers* (pp. 85–113). Chicago: University of Chicago Press.

Yinger, R. J., & Hendricks-Lee, M. S. (2000). The language of standards and teacher education reform. *Educational Policy, 14*(1), 94–96.

Zimpher, N. & Loadman, W. (1986). *A documentation and assessment system for student and program development*. Washington, DC: American Association of Colleges for Teacher Education.