An Organizing Framework for Using Evidence-Based Assessments To Improve Teaching and Learning in Teacher Education Programs[1]

By Gene E. Hall, Carol Smith, & Mary Beth Nowinski

There is widespread understanding of the need to evaluate teacher education programs. For example, the importance of conducting program evaluations has been addressed in past, as well as current, National Council for Accreditation of Teacher Education (NCATE) standards. There is a widespread expectation for teacher educators to provide evidence of effectiveness of regular, as well as innovative, programs. Additional impetus is present in the *No Child Left Behind Act* of 2001 with its mandate to school districts to place high quality teachers in every classroom. Policy makers, the media, fellow teacher educators, and teacher education candidates all assume that their programs are effective and that supporting documentation is readily available. Unfortunately, the history of teacher education program evaluation is spotty, evolutionary, and limited in scope. However, there exists a convergence of new expectations, policies, and methodologies for gath-

Gene E. Hall is a professor in the College of Education at the University of Nevada, Las Vegas; Carol Smith is vice president of professional issues and Mary Beth Nowinski is the technical assistance coordinator for the American Association for Colleges of Teacher Education, Washington, D.C.

ering, interpreting, and reporting evidence about program effectiveness and the quality of graduates.

There are a number of reasons for the current condition of teacher education program evaluation efforts. One under-estimated but significant factor is that teacher educators engage in intensive work with little time and attention available to conduct multi-year program level evaluation studies. A second factor is the direct costs of conducting systematic studies. There tend to be very limited resources for training observers and raters, conducting follow-up interviews, systematically observing teaching, and organizing databases. A third set of factors are related to the evolving views of what constitutes good evaluation. Forty years ago, program evaluations relied heavily on written questionnaires. Thirty years ago, national accreditation standards emphasized follow-up studies of graduates. Currently, some argue that the only study that will count is one that involves random assignment of subjects and a control group that receives no educational treatment. One of the authors of this paper argues that there should be no control groups in education. The control group required in the extreme scientific paradigm would receive no treatment, for example students in the literacy control group would not be taught to read.

Probably the most fundamental reason that program evaluations are limited is that there has not been a clear, consistent, and shared framework for organizing the many variables that comprise teacher education practice and relating these to evidence of effectiveness. Although we have been engaged in teacher education and program evaluation for a number of decades, only more recently have we and colleagues developed an organizing framework that others are finding useful. This framework was developed as part of the national Partnerships for Excellence in Teacher Education (PETE) Project, funded by the Ford and Carnegie Foundations, which is based at the American Association of Colleges of Teacher Education (AACTE). The framework and each of its components are described in this article.

Expectations for Program Quality

One of the key challenges in conducting Teacher Education Program (TEP) evaluations is determining the constructs and variables to be assessed. At different times and with each institution, developing agreement about the most critical elements has been arduous and the variables selected for study have been particular to each program. The prospective evaluator discovers that the meaning of quality in teacher education is a moving target. Different institutions advocate for different variables being most important and downplay the relative importance of others. Adding to the challenge, the shared paradigm for viewing teacher education and school quality has changed across time. The important indicators of quality in the 1970s were not the same as the quality indicators in the 1980s or the 1990s. Within each of these decades, different institutions advocated for certain variables, and the

more universally shared perspectives about quality in teacher education kept changing. For example, in October 1967, the U.S. Office of Education Bureau of Research issued a request for proposals to develop "Educational Specifications for a Comprehensive Undergraduate and Inservice Teacher Education Program for Elementary Teachers."

> Because of the key role that the teacher plays in facilitating learning, particularly with young children, he/she must have the most up-to-date theoretical and substantive knowledge and professional skills to perform successfully. To date, research and development activities have generated new knowledge, materials, and methodologies with great potential for improving the effectiveness and efficiency of the teaching-learning process. If funds are made available, institutions should be able at this time to completely restructure their teacher education program to include the best of what is known and available. (USOE, October 16, 1967)

In this case, teacher quality was defined as *up-to-date theoretical and substantive knowledge and professional skills*. There was the expectation that a systems analysis approach would be used in developing the specifications of the models. [Nine institutions received grants to design model programs. See the *Journal of Research and Development in Education,* Spring, 1969.] As a consequence, the developed models were based in task analyses of what teachers do, resulting in lists of skills and competencies that teachers should demonstrate. The skills were stated in behavioral terms and ranged in scope from narrow and specific to general.

By the 1980s there was a legacy of ten years of initiatives by many colleges and universities to develop innovative approaches to teacher education, and there were national networks of collaborating institutions. In addition to the programs to prepare elementary teachers, there were well-regarded secondary programs, such as the Nebraska University Secondary Teacher Education Program (NUSTEP) reported by Gades (1973), as well as five-year programs such as PROTEACH at the University of Florida reported by Smith (1984). Fuller and Bown (1975) reported on the development of a wide array of teacher education resources (module banks, protocol materials, and the Comprehensive Personal Assessment System) that were combined as the Personalized Teacher Education Program at the National R&D Center for Teacher Education at The University of Texas at Austin. What were missing were reports of program effectiveness.

In the late 1980s, a group of deans from colleges of education in research universities launched the Holmes Group and a new agenda for transforming teacher education, which was presented in *Tomorrow's Teachers* (1986):

> 1. To make the education of teachers intellectually more solid.
>
> 2. To recognize differences in teachers' knowledge, skill, and commitment, in their education certification, and work.
>
> 3. To create standards of entry to the profession — examinations and education requirements — that are professionally relevant and intellectually defensible.

4. To connect our own institutions to schools.

5. To make schools better places for teachers to work, and to learn. (p. 4)

While the Holmes Group advocated for more intellectual rigor and improving schools as professional work places, another set of deans and presidents from six state universities were forming the Renaissance Group where a core principle was for teacher education to be a campus-wide responsibility. The implied criteria for program evaluations were shifting. Should a program evaluator examine the intellectual depth of a program, or the extent to which faculty from all parts of the campus were engaged with teacher education? Or, should all of the above be evaluated?

The 1980s also was the time when classroom researchers developed a significant research base about teacher behaviors that were correlated with student achievement on standardized tests. The Direct Instruction teaching strategy was established and verified by scientific research (Brophy & Good, 1984). In the 1990s, researchers who used qualitative methods became significant contributors to the research base. At the same time, there was a philosophical shift where rational positivist perspectives were less valued and post modernist perspectives were more valued. Which perspective should be addressed in program evaluations? Should systematic observations in classrooms be replaced with ethnographic methods and examination of phenomenon that are not operationally defined?

National Standards for Program Accreditation

Even as education research and innovation in teacher education program design evolved, the expectations for national accreditation changed. In the 1980s, one of the NCATE standards was, "The unit maintains relationships with graduates from its professional education programs that include follow-up studies and assistance to beginning professionals." (p.8). In the last fifteen years, the NCATE standards have been revised three times. With each reformulation, an important new construct was introduced with implications for TEP evaluation studies.

Knowledge Bases for Professional Education (1992)

"The unit ensures that its professional education programs are based on essential knowledge, established and current research findings, and sound professional practice. . . . The faculty responsible for professional education collaborate in the design, delivery, and evaluation of curriculum for the unit's programs" (p.47).

Conceptual Framework(s) (1995)

"The unit has high quality professional education programs that are derived form a conceptual framework(s) that is knowledge-based, articulated, shared, coherent, consistent with the unit and/or institutional mission, and continuously evaluated" (p.15).

Assessment System (2000)

"The unit has an assessment system that collects and analyzes data on applicant qualifications, candidate and graduate performance, and unit operations to evaluate and improve the unit and its programs" (p.21).

To summarize, TEP evaluators are confronted with the ambiguities associated with significant changes in what is considered important. In the early 1990s, examining the knowledge base of faculty and courses would be important. By the late 1990s, evaluators were required to check on the overarching conceptual themes of the program and the extent to which teacher education students and partner schools were knowledgeable of these themes. Now, with the NCATE 2000 Standards, assessment of candidate learning is expected, along with establishment of an assessment system whereby the unit, faculty and candidates can trace how well each is accomplishing the goal of having candidates become quality teachers.

Trends in Schools

Another set of factors that influence TEP evaluations are the practices and needs of today's schools. Schools have turned to performance assessment as the method for linking student learning to what teachers teach. This movement also is directly related to the significant emphasis on standards as statements of expectations for student learning. National, state and school district standards for student learning represent year-long and multi-year road maps. Each teacher needs to have tools to monitor student benchmarks that represent the smaller steps on the way to achieving yearly standards. Performance assessments have become the strategy (Wiggins, 1998; Stiggins, 1998; Asp, 2000).

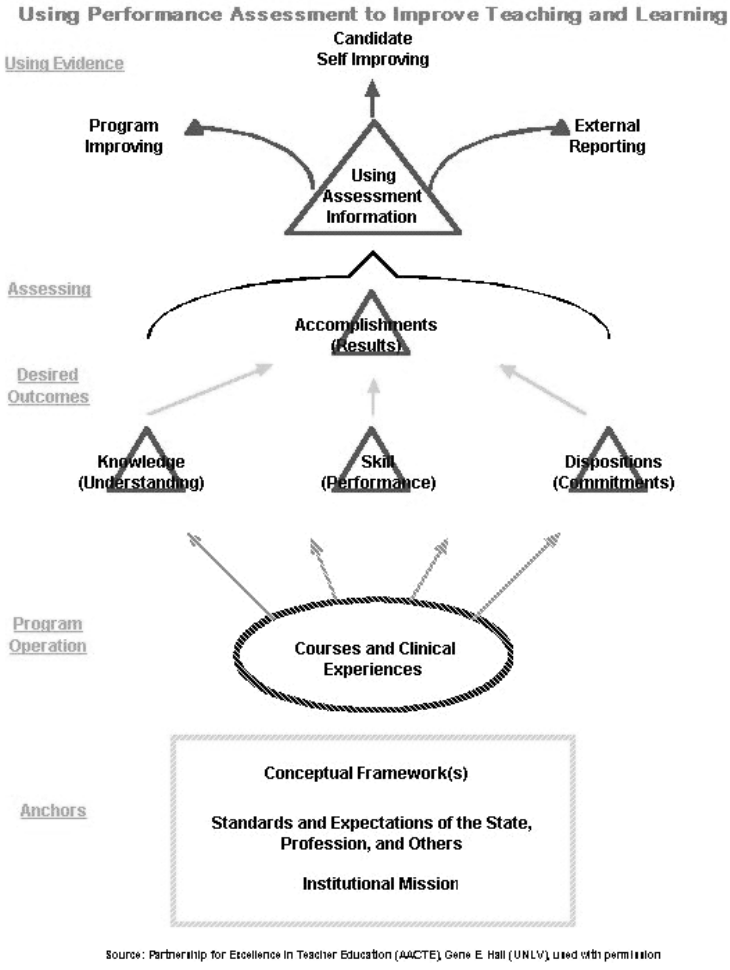
An Organizing Framework for Assessing Innovation in Teacher Education

Teacher educators must make a major paradigm shift in thinking that is represented in accreditation standards, the standards-based expectations for student learning, and the increasing use of performance assessment methods in schools. They have to stop thinking in terms of *Inputs*, "What I am teaching," and start thinking in terms of *Outputs*, "What are my students learning?" This shift in thinking is reflected in the NCATE 2000 Standards as well as all of the K-12 content standards with the diminished attention to evaluation of course syllabi, the number of books in the library, and the amount of technology. NCATE standards now ask for evidence that candidates are learning as they progress through the teacher education program and whether or not the graduates of programs, first year teachers, can make a difference in the academic success of the students they teach.

Making the paradigm shift from thinking in terms of inputs to thinking in terms of outputs along with using systematic evidence are important elements of the AACTE

Partnerships for Excellence in Teacher Education (PETE) project. The purpose of PETE is to provide technical assistance to institutions that are preparing for their initial NCATE accreditation. PETE has some fifty consultants who work with institutions during the three to four years that it typically takes to prepare for and successfully complete the initial accreditation process. With the emergence of the NCATE 2000 standards, the Steering Committee for PETE developed a picture to portray graphically how all of the various pieces and processes fit together. The outcome of this effort is the organizing framework presented in Figure 1. It is easiest to understand this framework when it is discussed one layer at a time, starting at the base.

Figure 1



Source: Partnership for Excellence in Teacher Education (AACTE), Gene E. Hall (UNLV), used with permission

Anchors

The philosophical and conceptual underpinnings of teacher education are derived from a number of sources. One important source is the *Institutional Mission*, which should reflect that teacher education has an important role across the campus. The various statements of *Standards and Expectations* are another program anchor. The national P-12 content standards, as well as state and local school district statements of expectations for student learning, are important considerations in developing a teacher education program. There are other standards that should be considered such as those of INTASC and the National Board for Professional Teaching Standards. Bringing all of these expectations together takes place in the process of developing a Conceptual Framework

The Conceptual Framework is the place where the philosophical positions of faculty and their beliefs about how candidates and students learn become an integrated whole. One of the other agreed upon features about teacher education that has become well established in the last thirty years is that a *program* is more than a collection of courses and field experiences. This is especially true when teacher education is studied from the point of view of candidates as they experience the program. The Conceptual Framework is the conceptual and philosophical underpinning where standards, institutional mission, and expectations of the profession and society come together into a statement of the whole. What is valued, what is unique, and what is important in teacher education at this institution should be clear and offered with passion in the statement of the Conceptual Framework. Evaluation studies need to be grounded in these anchors.

Program Operation

Nearly all teacher education programs are organized into *Courses and Clinical Experiences*. As obvious as this is, what is less often made clear is the reasoning behind the particular content, processes and sequences that make up the program. Very little should be offered that is not directly linked back to the Anchors and clearly rationalized through the Conceptual Framework. The tasks of searching for, inventorying, and appraising the rationales for the hypothesized value of each course and clinical experience can be facilitated by program evaluation. For example, a mathematics methods course that requires candidates to study the mistakes students make when learning algebra would be supported by the NCTM standards. An evaluation study could assess the performance of candidates as they examine student work and plan the next steps of instruction.

Desired Outcomes

Four important domains of desired outcomes have been identified and are shown in Figure 1. (The meaning of the triangles which appear over each domain

will be explained subsequently.) *Knowledge (Understanding)* is a necessary indicator of teacher quality. Teachers must know and understand the content of the subject(s) they teach. They must know and understand the foundations of education including educational psychology, sociology, philosophy and history. The debate is about how much of each of these subjects teachers must know.

Future teachers must develop *Skill (Performance)* in teaching. A comment often heard refers to the teacher who "really knows the subject, but can't teach it so that I can understand it." Knowing the subject is one thing, explaining it in understandable ways in the classroom is quite another. Skill in teaching includes managing instruction, keeping students safe, and making lessons interesting. Skill in teaching also means effectively teaching in classrooms with culturally diverse students, English Language Learners, and students with special needs. Assessing student learning is another important skill, especially because so much assessing is conducted during the act of teaching.

In the last decade, increasing attention has been directed to *Dispositions (Commitments)*. Teacher enthusiasm about and confidence in their knowledge of the subjects they teach is related to developing student interest and engagement. A critical component of dispositions is the ways that teachers attend to and treat boys versus girls, children of different colors, those with special needs as well as students who are English Language Learners. Other key dispositions include commitment to continually improving oneself as a teacher and supporting or leading within the profession.

Accomplishments (Results) represent a focus that is becoming more important, and we believe will become central in the near future. The most critical outcome is what happens with the students. The earlier studies by Brophy and Good (1984) as well as the more recent studies of Sanders (1998) make it clear that some teachers are more effective in helping students learn. They achieve results. Today's teacher education programs must clarify the expectations for candidates in terms of the differences they should make in the students they teach. In each and every field experience, candidates should be expected to prepare some form(s) of evidence that shows the effects of their teaching efforts. *Teacher Work Sample* (Girod, 2002) is one method for structuring teaching tasks in ways that allow candidates to examine the learning of their students.

Assessing

Each of the triangles over the outcome domains shown in Figure 1 represents an area where assessing candidate performance should occur. Note that the triangles are placed over the four domains of desired outcomes, and not over the program operation or anchors so as to emphasize the paradigm shift from thinking in terms of inputs to thinking in terms of outputs. The assessment work is targeted on candidate learning and performance. Granted, there will be links back to program

operation and anchors, but the primary work of assessment must be centered on desired outcomes.

There are clear and direct links between Assessing, Program Operation, and Anchors. For example, the rationale and derivation of Desired Outcomes should be found in the Conceptual Framework and there should be courses and/or clinical experiences designed to help candidates learn the desired outcome. Symbolically the starting point needs to be expectations for the difference candidates will make in the classroom, i.e., Accomplishments/Results, which can be traced to Assessing of Outcomes, Program Operation and the Anchors, especially as developed in the Conceptual Framework. As suggested by Elliot (2003), teacher educators can choose from a wide variety of assessment methods.

Assessing Knowledge and Understanding

While the foundational knowledge required for effective teaching extends well beyond content knowledge in the licensure field, this narrower concept has become the primary driving force for assessment of teachers. Although few teacher educators regard the typical state licensure test as an adequate measure of content knowledge of teaching, programs tend to cede the assessment of comprehensive knowledge to these standardized tests and otherwise rely on measures within individual courses.

The most significant recent developmental work by teacher education faculty appears to be in sampling specific areas of content knowledge through performance-based measures that produce documentation for portfolios or other program assessment structures. For example, portfolio assessments in teacher education pioneered by Alverno College test knowledge within the framework of competencies identified for their candidates. Assessment of teacher knowledge developed by Missouri Western State College relies on a generic rubric that identifies developmental levels of candidates' intellectual understanding of P-12 students' engagement with content. In another example, a self-assessment instrument developed by Johns Hopkins University to gauge candidates' content knowledge in alternate route programs relies on faculty identification of content domains linked to the state's P-12 academic standards.

Standards for assessment generally and those of NCATE for teacher education call for multiple measures of such dimensions as candidates' content knowledge and understandings. The INTASC design for teacher assessment similarly envisions portfolio evidence and some performance measures as ways of sampling specific areas of content knowledge with complementary assessments to gauge adequate breadth of knowledge in related area(s).

Assessing Skill and Performance

Models developed outside teacher preparation programs have had some of the most profound impact on how faculty think about assessing candidates' skill and performance. For example, the professional assessment model developed by the

National Board for Professional Teaching Standards (NBPTS) has become the archetype for testing skill and performance. This approach combines portfolio documentation, including video, with on-demand assessments in which performance tasks are linked to specific assessment prompts. The state licensure model developed by Connecticut through its BEST (Beginning Educator Support and Training) program combines portfolio assessment with beginning teacher support in an approach that has a strong focus on content knowledge as well as teaching skills. Both of these assessment models address teaching skills in the context of applied knowledge of content, pedagogy, and relevant developmental aspects of individual students.

Assessment of skills and performance presents multiple challenges for teacher education faculty, including assurance of reliability and validity. As noted in Dunkin's (1997) review of evolving approaches to assessing teachers, "the validity of the emergent methods when they are combined into programs of evaluation seems seldom to have been scrutinized on a large scale" (p. 49). Assessment credibility in the NCATE 2000 standards is addressed through "fairness, consistency, accuracy and avoidance of bias" (p. 23). Accreditation reviews under these performance-based accreditation standards are identifying aspects of assessment system design and implementation as major areas in need of improvement for many teacher education programs. Experiences of institutional networks such as the Standards-based Teacher Education Project (STEP) indicate that another major area for development is greater involvement of arts and sciences faculty in assessing candidates' applied content knowledge and related performance evidence.

Assessing Dispositions

Professional dispositions are explicitly incorporated in such frameworks as the INTASC licensure standards and state program accreditation standards. For example, Indiana's teacher licensing standards incorporate content-specific expectations for dispositions related to each area of subject matter content in which that state's teachers are licensed. While national and state standards set the context for required assessment of dispositions, much of the consensus as to which dispositions are identified and how related evidence is assessed proceeds on a program-by-program basis, usually within the context of the institution's particular conceptual framework. For example, the University of Indianapolis requires an essay and interview for admission to teacher preparation in which practicing teachers participate as assessors, the University of Memphis has a comprehensive structure to articulate and assess dispositions throughout their teacher preparation program.

Assessment of dispositions is a complex process that requires development of explicit statements and criteria, opportunities within the program to assess dispositions, and instruments or processes that are appropriate to the factor being assessed and to the high-stakes nature of the decisions being made about candidates. While statements of expectations for dispositions may indicate areas of considerable

overlap and consistency, there has not yet been an effort to gain national consensus on either the dispositional factors or the assessment instruments used to assess them. Whether such consensus is appropriate or feasible appears to be an open question.

Assessing Accomplishments and Results

Assessing results in teacher preparation has moved from using descriptive evidence of process and instructional activity to using more objective evidence of teacher knowledge, skills, and dispositions and is rapidly moving to encompass the full continuum of evidence of impact on P-12 student learning and links to teacher performance. Many teacher educators seem to resist acceptance of student learning as evidence of program effectiveness (see Millman, 1997). Instead, they seem to be moving toward a definition that focuses on candidates' ability to build appropriate assessments, reflect on data gained from assessments, diagnose individual and group needs, and determine next steps in adapting instruction to meet student needs. The best known example of focusing directly on student learning as part of candidate assessment is Western Oregon University's model of teacher work sampling (Girod, 2002) and adaptations developed by Western Kentucky University and an affiliated network of institutions in the Renaissance Group (Denner, Salzman, & Harris, 2002).

Challenges to having student learning as a significant component of candidate evaluation include the limited amount of time in which most candidates have responsibility for lesson planning, development, and delivery as part of their preservice field experiences, and the difficult patterns of P-12 standards and assessments that exist in many states and school districts. A larger conceptual challenge that faculty face as they design credible approaches to the student learning issue is how to make the distinction between research-based evidence that a given teacher can claim impact on specific student learning gains in a causative sense, as opposed to alternative frameworks for describing specific linkages between teacher knowledge/skills/instruction practices and the ability of students to reach learning goals.

Researchers who study change processes have brought additional conceptual and methodological clarity to what is necessary in assessing Accomplishments and Results. As Hall (1999) has described, there is a chasm between current teacher practices and increasing test scores. If teachers and schools continue as they have in the past, there is no reason to expect student learning to increase. Simply establishing another new education accountability policy, or purchasing a school reform approach, will not by itself result in improvements of student learning. The necessary prerequisite is that teachers *change* by implementing the new practices. Instead of assuming that teachers and schools can make the Giant Leap across the chasm, there needs to be an *Implementation Bridge* to facilitate teachers making change. George, Hall, and Uchiyama (2000) documented that as teachers move across the bridge there can be increases in student learning. One implication of this

research for teacher education is that rather than simply examining student test scores, evaluation designs should first assess the extent to which candidates have implemented best practices (Hall & Hord, 2001).

The Search for Good Measures

One of the consequences of the assessment movement is increasing teacher educator interest in finding measures that match well with the Anchors for their program, are practical to use, and that are credible. One wide-spread approach is to continue using current measures such as course exams and student teacher evaluation forms. Unfortunately, most of these have little grounding in program conceptual frameworks, typically have no estimates of reliability and validity, and are likely to not be performance based. Two other approaches are to develop new measures and to search out established measures. There are two national efforts to review established measures. In both cases, the efforts are screening for quality and relevance.

Elliott (2003) has led the NCATE Assessments Example Project which collects examples of assessments submitted by teacher educators. From these examples, criteria for good assessments were derived. As Elliott (2003) wrote:

> Faculty want to be responsive to NCATE's insistence on assessment results as evidence that candidates have mastered unit and program standards prepared by NCATE and its affiliated program specialty associations. Moreover, they are trying to build on the best that we know about use of assessments for learning - such as "alignment" of assessments with standards, use of multiple measures, and evaluating a wide range of knowledge, skills, and dispositions. Faculty who are searching for examples expect that someone, somewhere, in professional educator preparation programs has developed assessments that can serve to demonstrate what is possible. What is needed is not an elegant new design from external testing experts, or the latest research efforts in test development. Instead, what is needed are assessments created by colleagues in other institutions, and actually used in courses and experiences of professional educator preparation programs. That way we can know that the examples are not out of reach, but are practical for adaptation and as sources for new ideas. (pg. vii)

The second initiative to review and compile measures is being conducted by Raths (2003) for the Teacher Education Accreditation Council (TEAC). Raths reviewed a wide array of measures with close attention to issues of reliability and validity as well as the research literature on quality measures that can be relevant to teacher education.

Using Evidence

In the past, once assessment information was collected, compiled and reported, little was done with the information. One way to use assessment information is in *Program Improving*. If elementary student teachers are not feeling confident (a disposition) in teaching mathematics, this information should lead to changes in the methods course and perhaps the content knowledge requirements. If the analysis

of student teaching work samples indicates that candidates need to know more about the construction of assessment tasks and how to base them in standards and benchmarks (Knowledge and Skills). Evidence collected from program evaluation should be used to inform and focus efforts to improve the program.

Assessment information should also be used for *Candidate Self-Improving*. Data from peer observations, candidate's team teaching small groups of students, and clinical supervisor feedback should be seen by candidates as important information to guide self improvements rather than just as the basis for final grades. In these examples, Dispositions can be as important as the structure of a particular teacher education experience. An important implication of using evidence is that the program evaluation data can be made available to candidates in systematic ways. For example, the assessment system at State University of New York Cortland relies heavily on candidates entering their own data and monitoring their accumulating performance records.

Increasingly assessment information is necessary for *External Reporting*. Accreditation reports, program reviews, and the public want to know more about program effectivensss and the quality of graduates. It is likely that over the next several years there will be more regulations for federal reporting that require particular kinds of evidence of candidate quality and program effectiveness. Institutions must now provide evidence of effectiveness for Title II programs, state accreditation, and national accreditation. All of these reports require evaluation studies that are planned, prioritized and ongoing.

In summary, having an organized and systematic set of assessment activities, in other words an assessment system, has become an essential component for all teacher education programs. What is being proposed in this article is that assessment work needs to be grounded in the Anchors, Program Operation and Desired Outcomes. When all of the pieces and areas of activity that are required to do quality teacher education are thought about separately the burden can be overwhelming; which is why we developed the organizing framework that is presented in Figure 1. We too needed a picture that mapped the critical components and their relationships in order to maintain clarity about what is necessary to do when teacher education is done well.

A consequence of developing a comprehensive picture is that there is increased clarity about how program evaluation should be conducted. TEP evaluations must focus on assessing candidate performance in four domains: (1) Knowledge/Understanding, (2) Skill/Performance, (3) Dispositions/Commitments, and (4) Accomplishments/Results. All other evaluation activity needs to spin off this target. Holding this focus facilitates making the necessary paradigm shift from thinking in terms of Inputs (Anchors and Program Operation) to Outputs (Desired Outcomes).

Looping of evidence also becomes possible with such a holistic view. For example, when evaluations are designed with links back to program operations and the conceptual framework, the study findings can provide feedback to faculty who

can then make changes in courses and clinical experiences. The findings can be shared with candidates in order to empower them to be data-based in their own self-assessments and reflections. Triangulation of findings becomes possible.

Conclusions

Three major themes have been developed in this paper. First, teacher education has an extensive record of program innovation and related knowledge bases. At the same time, there has been constantly changing criteria for judging the effectiveness of programs and the quality of the teachers produced. Second, there is a need for an umbrella picture, an organizing framework, for viewing the foundation components, processes, and outcomes of teacher education. At the center of such a view has to be teacher candidate and student learning, which represents a paradigm shift from thinking in terms of inputs about course and experiences, to thinking explicitly about outputs. Third, assessing candidate learning and evaluating programs needs to be based in evidence of candidate and student learning.

The framework presented here offers the bonus of being able to compare TEP evaluation findings from different programs and institutions. If there can be a shared effort within and across programs then it will be possible to provide strong evidence of program effectiveness and increasing candidate quality, which are the expected outputs from the complex and rich process that is simply called teacher education.

Note

[1] The authors wish to acknowledge the funding support of the Ford and Carnegie Foundations for the activities of the AACTE PETE project, which provided the field experiences and institutional contexts that led to the development of the framework presented in this manuscript.

References

Ascher, C. (1993). *Changing schools for urban students: The school development program, accelerated schools, and success for all. Trends and issues No. 18*. New York: ERIC Clearinghouse on Urban Education, Teachers College, Columbia University.

Asp, E. (2000). *Using teacher judgment to assess student achievement*. Castle Rock, CO: Douglas County School District.

Brophy, J. E., & Good, T. L. (1984). Teacher behavior and student achievement. Occasional Paper No. 73. *Handbook of research on teaching (3rd ed.)*. East Lansing, MI: Institute for Research on Teaching, College of Education.

Denner, P, R., Salzman, S, A., & Harris, L,B. (2002, February). *Teacher Work Sample Assessment: An accountability method that moves beyond teacher testing to the impact of teacher performance on student learning*, A paper presented at AACTE Annual Meeting, (New York, NY).

Dunkin, M.J. (1997). Assessing teachers' effectiveness. *Issues in Educational Research*, *8*(1), 37-51.

Elliott, E, J.. (2003). *Assessing education candidate performance: A look at changing practices.* Washington, DC: National Council for Accreditation of Teacher Education.

Fuller, F. F., & Bown, O H. (1975). Becoming a teacher. *Seventy-fourth Yearbook of the National Society for the Study of Education* (pp. 25-52). Chicago, IL: The National Society for the Study of Education.

Gades, R. E. (1973). NUSTEP Nebraska University Secondary Teacher Education Program. *Journal of Business Education*, *49*(3), 119-20.

George, A. A., Hall, G. E., & Uchiyama, K.. (2000). Extent of implementation of a standards-based approach to teaching mathematics and student outcomes. *Journal of Classroom Interaction*, *35*(1), 8-25.

Girod, G. R. (2002). *Connecting teaching and learning: A handbook for teacher educators on teacher work sample methodology.* Washington, DC: American Association of Colleges of Teacher Education.

Hall, G. E. (1999). Using constructs and techniques from research to facilitate and assess implementation of an innovative mathematics curriculum. *Journal of Classroom Interaction*, *34*(1), 1-8.

Hall, G, E., & Hord, S.M. (2001). *Implementing change: Patterns, principles and potholes*. Boston: Allyn & Bacon.

Millman, J. (Ed.). (1997). *Grading teachers, grading schools: Is student achievement a valid evaluation measure?* Thousand Oaks, CA: Corwin Press.

National Council for Accreditation of Teacher Education. (1987). *Standards, procedures, and policies for the accreditation of professional education units* (p. 42). Washington, DC: National Council for Accreditation of Teacher Education.

National Council for Accreditation of Teacher Education. (1992). *Standards, procedures, and policies for the accreditation of professional education units* (p. 47). Washington, DC: National Council for Accreditation of Teacher Education.

National Council for Accreditation of Teacher Education. (1997). *Standards, procedures, and policies for the accreditation of professional education units* (p. 15). Washington, DC: National Council for Accreditation of Teacher Education.

National Council for Accreditation of Teacher Education. (2002). Professional *Standards for the accreditation of schools, colleges, and departments of education* (p. 21). Washington, DC: National Council for Accreditation of Teacher Education.

No Child Left Behind Act of 2001. (H.R. 1), 107 Cong., 110 (2002) Enacted.

Raths, J, (2003). *A compendium of measures for teacher education.* Newark, DE: Teacher Education Accreditation Council.

Sanders, W, L. (1998, December). Value-added assessment. *School Administrator, 11*(55), 24-27.

Smith, D, C. (1984). PROTEACH: An extensive preservice teacher preparation program. Gainesville, FL: University of Florida. ERIC No. E0250317.

Stiggins, R. J. (1998). Classroom assessment for student success. *Student Assessment Series.* National Education Association. Washington, DC.

The Holmes Group. (1986). *Tomorrow's teachers: A report of the Holmes group.* East Lansing, MI: Michigan State University, Author.

Wiggins, G. (1998). *Educative assessment: Designing assessments to inform and improve student performance.* San Francisco: Jossey-Bass Publishers.