

Exploring Existing Teacher Assessment Practices in California: Implications for Potential Reforms

**By Jo Ann Izu, Claudia Long,
Kendyll Stansbury, & Dennis S. Tierney**

Jo Ann Izu, Claudia Long, Kendyll Stansbury, and Dennis S. Tierney were staff members at Far West Laboratory for Educational Research and Development, San Francisco, California, when the research described in this article was undertaken. Izu remains a staff member at Far West Laboratory. Long is currently an independent research consultant. Stansbury is now a staff member with the Connecticut Department of Education, Hartford, Connecticut. Tierney is now a consultant with the Professional Services Division of the California Commission on Teacher Credentialing, Sacramento, California.

Senate Bill 1422 (Bergeson, 1992) provides an unusually open-ended opportunity for all individuals interested in improving the preparation of classroom teachers in California. It calls upon the state's independent teaching standards board, the Commission on Teacher Credentialing (CTC), to "review the requirements for earning and renewing multiple and single subject (elementary and secondary) teaching credentials..." This statute allows significant latitude in defining the purposes, scope and issues to be studied as a part of the review. The CTC is interpreting this legislation as an opportunity to "invent a new system of teacher credentialing for 21st century schools" (CTC, November, 1994).

Teacher Assessment

This landmark legislation grew out of earlier efforts in California to address some long-standing concerns about teaching as a profession. Beginning with the California Commission on the Teaching Profession, known as the “Commons Commission Report,” in 1985, the state attempted to develop new policies to improve the teaching profession and thereby make the public schools better. This report, and others that followed, called for a restructuring of the teaching career and the establishment of more rigorous standards. In 1988, the Bergeson Act called for “new policies to govern the support and assessment of beginning teachers,” launching the California New Teacher Project (CNTP). The CNTP was a large scale, four-year pilot test of new approaches to supporting beginning teachers in the first two years of professional work, coupled with an evaluation of a wide array of teacher assessments intended to improve their practice through more detailed and meaningful feedback.

The Far West Laboratory for Educational Research and Development (FWL) was awarded a contract to conduct the research related to teacher assessments. This article reports the essential findings of one phase of the assessment studies conducted during the CNTP and suggests some implications from that study for the impending reform efforts embodied in the SB 1422 study. Additional research conducted under the auspices of the Beginning Teacher Support and Assessment (BTSA) project are summarized in accompanying articles in this special edition of *Teacher Education Quarterly* and in other publications.

The final phase of the FWL assessment studies examined the existing teacher assessment practices in California from the point of entry to a teacher education program through the awarding of tenure by a school district. Policymakers wanted to know more about the state of teacher assessment practices before launching new policy initiatives.

This article limits itself to the findings from that last phase of the CNTP research and relevant implications for future efforts at modifying the assessment practices across institutions and during the early years of a teacher’s professional development.

CNTP Assessment Studies

As part of the CNTP, co-sponsored by the CTC and the California Department of Education (CDE), FWL conducted a study of existing teacher assessment practices in California during the spring and early summer of 1991. The study focused on the evaluation of prospective and beginning teachers from undergraduate content area coursework through the awarding of tenure by a district.

The research design called for in-depth examinations of assessment practices in a small number of organizations which collectively represented the diverse contexts in California and at the same time maximized the diversity of assessment practices. Documents were examined and staff interviewed from 12 institutions of

higher education (IHEs), 12 districts (LEAs), and the CTC. Each different type of organization constituted a separate sub-study. In addition, 22 beginning teachers were interviewed about their experiences, both to provide an otherwise missing perspective and to see how the totality of assessment practices did and did not cohere. With the exception of the statewide examinations accepted or required by the CTC, each sub-study included teachers, campuses, and districts representing rural, suburban, and urban contexts, with size ranging from small to very large. The IHEs included campuses from the California State University (CSU) and University of California systems as well as private and independent colleges and universities. Some organizations were selected because they were known to have implemented innovative assessment practices. Beginning teachers represented a range of performance on various assessment instruments pilot tested in another phase of the CNTP.

In addition to examining relevant documents, the researchers interviewed at least one person in every role group that participated in key decision-making with respect to teacher evaluation. Administrators provided overviews of each assessment point for which they had responsibility. All respondents, including the beginning teachers, were asked to describe the processes in which they were directly involved. Twenty-one documents, on average, per IHE and up to 15 per LEA were examined. Interviews with 172 IHE staff and 152 LEA staff across the 24 organizations were conducted by trained interviewers.

Information was collected on eight points at which beginning teachers are often examined in their teaching career, including:

- 1) undergraduate coursework examinations in content areas;
- 2) admission to teacher preparation programs;
- 3) advancement to student teaching;
- 4) performance assessment in student teaching and other teacher education courses;
- 5) CBEST and other state-mandated examinations to satisfy credentialing requirements;
- 6) applications for employment;
- 7) formative assessment of new teachers for support purposes;
- 8) summative assessment of new teachers for continued employment decisions.

Site summary forms were completed for each IHE and LEA studied. The summaries highlighted both common and unusual policies and site-level practices. After writing summaries of findings for each separate sub-study, the four authors met and developed overall findings across sub-studies, which are described in the remainder of this article.

Focus of Assessments

Subject matter knowledge, which provides a foundation for the development of skills in content pedagogy, is the primary focus of the assessments for entry into

Teacher Assessment

a teacher preparation program. Assessment generally occurs through either a standardized examination approved by the state, or examinations in courses prescribed by a subject matter preparation program. Subject matter knowledge seldom gets re-examined at later points of assessment, as both IHEs and LEAs rely heavily on this initial screen. Subject matter knowledge may also be indirectly evaluated when IHEs and LEAs look at content pedagogy.

Current methods of assessing subject matter knowledge generally focus on breadth, not depth, of knowledge. They miss such deficiencies as inadequate knowledge of the structure of the discipline, an inability to sequence topics in a meaningful way, and a lack of ability to apply basic constructs (*e.g.*, fractions, different stages of the writing process). However, the revised subject matter tests recently adopted by the CTC include elements of application designed to address the last deficiency.

General pedagogy, especially classroom management, is the major focus of all assessments from student teaching through the early years of teaching. The state relies on IHE assessments of general pedagogy for credentialing requirements. Looking across the assessments conducted by the IHEs, LEAs, and the CTC, evaluation of skills in content pedagogy are consistently lacking in depth. During student teaching, content pedagogy is minimally assessed through observations and examination of lesson plans, and standards tend to be relatively low. The low standards do not seem to be grounded in a clearly defined perspective on teacher development. As with general pedagogy, the CTC again relies on IHEs to evaluate content pedagogy for credentialing purposes. Content pedagogy was also often minimally assessed by the LEAs in our sample at the point of hiring and/or in the first year of teaching. Skill in content pedagogy is most often defined by the LEAs as the ability to use the instructional techniques emphasized by the district.

Knowledge of students and of teaching diverse students is generally a very minor focus of all the assessments from student teaching on. Only a small amount of feedback during student teaching is focused on teaching diverse students, and tends to be directed toward classroom management issues or teaching students of differing ability levels. Knowledge of students is most directly assessed by LEAs at the point of hiring, when specific questions are sometimes asked about a teacher's familiarity or experience with certain types of students.

General academic ability and personal attributes are always assessed. General academic skills are evaluated at every entry point, usually through multiple measures: CBEST, GPA, and occasionally through performance assessments (*e.g.*, a writing sample). Personal attributes are primarily assessed through observation and interviews. Although IHEs and LEAs differ in some of the specific attributes they emphasize, almost all look for enthusiasm, rapport with children, and a commitment to teaching.

Forms of Assessments

Observation is the primary form of assessment used by IHEs and LEAs to assess a beginning teacher's knowledge, skills, and abilities during student teaching and the first years of teaching. Document review is commonly used at points of entry; however, IHEs and LEAs vary in how thoroughly they review materials and in the inferences they make from the same information. Written examinations are used by IHEs and the CTC to assess subject matter knowledge and basic skills proficiencies. IHEs also use written examinations to assess general and content knowledge, usually before the state of student teaching.

Alternative performance assessments (*e.g.*, unit/lesson plans, portfolios) are secondary forms of assessments sometimes used from teacher education coursework on. The CTCn has incorporated performance assessment into the subject matter knowledge examinations. Interviews are commonly conducted by IHEs at admission to the teacher preparation program and by LEAs at hiring. Alternative methods of assessment such as student achievement data and student evaluations of the teacher are occasionally used during student teaching and the first years of teaching.

Technical Quality

The technical quality of these assessment practices varies considerably both between and within institutions, but with the exception of the standardized state assessments, it is generally low. Clearly stated expectations and performance standards are rarely a characteristic of present assessments. Moreover, the assessments with the clearest criteria and standards are those that measure general academic ability and basic skills proficiencies; they do not measure teaching skills. Criteria used in present assessments of teaching skills are most often represented by broad categories whose interpretation is largely left up to individual assessors. Rating scales, when used, are rarely defined. In some small credential programs and districts, frequent conversations among the limited number of assessors increases the uniformity in interpretation and rating of criteria. In larger programs and districts, extensive training would be required to achieve the same result.

Not surprisingly, preparation of assessors, when it occurs, is largely oriented more to procedure than to substance. Even this extent of preparation varies by assessment point. For instance, training provided by districts for assessments for retention is common, while training for hiring assessments is not. Although beginning teachers interviewed said that most of the detailed and frequent feedback during student teaching is from cooperating (or master) teachers, the cooperating teachers themselves tend to receive no training for this role. Over one-third of the cooperating teachers interviewed reported some confusion over the meaning of some parts of the form they use to rate student teachers.

An Assessment System?

Desired characteristics of an ideal statewide assessment system are: rigor, thoroughness, consistency, and helpfulness. The extent to which these constructs characterize the assessment practices identified in the overall study is discussed separately for each construct.

Rigor

Both IHE and LEA assessment practices are uneven in terms of rigor, both across institutions and within institutions and credential programs. Evaluation criteria are identified and sometimes defined by detailed examples. However, standards, *i.e.*, the rating categories used, are rarely defined. Those IHEs and LEAs that define the constructs being assessed, either conceptually or by providing examples, usually have the most rigorous assessments as well.

Few documents examined reflected a well-articulated perspective on the development of teaching skills. The form and procedures for evaluation of teachers at an early stage (*e.g.*, student teacher in first placement, first-year teacher) were almost always the same as that for teachers at later stages (*e.g.*, student teacher in second placement, 20-year veteran). Assessors reported taking the extent of a teacher's experience into account in their evaluation, but the basis for doing so was not clear, and in many cases appeared to vary with the individual assessor.

LEAs differ in their capacity to maintain high standards in the evaluation of prospective and beginning teachers. Standards for new teacher assessment by LEAs are driven by market forces, where beginning teachers are judged relative to those who might be expected to replace them. Districts vary in their attractiveness to teachers, and some of the most influential negative factors (*e.g.*, inner city location, per-pupil funding levels) are beyond a district's control. Although recruitment strategies have some impact, a district's attractiveness strongly influences the number and quality of applicants. As a result, districts with many high quality applicants are able to apply high standards, while districts with difficulty recruiting credentialed teachers are not. However, districts with severe staffing problems do employ standards by which they reject credentialed teachers. These standards often include more emphasis on knowledge of and empathy with the type of students in the district and less emphasis on pedagogy, reflecting a belief that pedagogy is more easily taught.

In terms of identifying candidates who should be eliminated from the profession, the assessment is most rigorous at the entry points. From the perspective of the beginning teachers interviewed, the rigor of assessment depends on the variety of areas being assessed, the amount of work expected, and the level of expectations held by the assessor(s). Most considered their student teaching assessments to be more rigorous than those experienced during the first years of teaching.

The rigor of assessments was affected both by resources and available

methodology. State assessments exhibit high technical quality, in part because the high cost of well-developed assessments can be spread across a large number of candidates, and a source of income is available to fund administration, development, and implementation costs.

For eliminating candidates from the system, IHEs and LEAs tend to favor assessments that are “appeal-proof,” which tends to mean one of two characteristics. The first characteristic is that the assessment criteria are directly measurable. For example, while the validity of using a GPA as a measure of academic skills may be argued, whether or not a specific GPA meets the standard cannot. However, whether or not a teacher is using an appropriate instructional technique is less clear, and more vulnerable to conflicting interpretations. The ease with which indicators of skills can be measured varies across teaching domains. Classroom management is relatively easy to observe and document, while content pedagogy and knowledge of students are not.

The second, and related, characteristic of an “appeal-proof” assessment is the degree to which there is a consensus on a definition of important teaching skills across contexts and teaching styles. One reason that content pedagogy and knowledge of students are difficult to document is that they are difficult to define across different teaching contexts and styles.

LEAs are not required to give reasons for releasing teachers prior to receipt of tenure, and assessors in many of the sample districts stated that they did not explain decisions not to re-hire beginning teachers to avoid potential legal problems. Some districts even have recommended guidelines based on legal criteria as to what assessors should and should not tell teachers at the point of dismissal.

Thoroughness

Assessments seem to be conducted without the benefit of a general teaching framework which serves as a context in which to discuss a beginning teacher’s knowledge, skills, and abilities. Often there are lists of competencies, but is not clear how the competencies interrelate. Some items seem to be more characteristic of experienced than student or beginning teachers (*e.g.*, for a student teacher, “paces individual lessons appropriately”). Without a general teaching framework, each teacher is left to generalize from their particular teaching experiences and invent their own framework. Prospective and beginning teachers seem to have difficulty in putting feedback about specific actions and reactions into a larger framework so that they see alternative ways of approaching problems. The lack of a framework also facilitates the rejection of negative feedback, because the teacher is unable to link that specific feedback to a larger principle or issue. Thus, negative feedback is more vulnerable to being ignored, or even rejected due to being attributed to a personality conflict or lack of contextual understanding by the assessor.

Information from previous assessments is rarely used for diagnostic purposes, both across institutions (*e.g.*, from IHE to LEA) and within institutions (*e.g.*,

Teacher Assessment

information from education coursework used in student teaching; formative and summative evaluations of new teachers). Where information from the IHE is available to the LEA for hiring of new staff, a lack of time to consider the information means that it most often either is not used or is used only in the final stages of the selection process, where the number of active candidates is considerably reduced. Even if assessors or institutions are willing to share and use information, there are legal liabilities with respect to negative information which affects the information sent. Generally, only “fatal” weaknesses of applicants are disclosed. Minor to major weaknesses are often identified by omission, requiring an evaluator to go well beyond what is written in letters of recommendation. Typical comments were: “Unless the letter of recommendation specifically states that classroom management is strong, we assume that it isn’t,” and “Listen to what they don’t say as much as what they do say.”

Fairness

The absence of a conceptual framework of teaching makes it difficult to set fair levels of expectations. Because of the lack of explication of assessment standards, fairness often depends on the professional judgment of people who are not always well trained. The differences in teaching contexts, especially for IHEs, also contribute to the difficulty of maintaining fairness, perhaps contributing to the reluctance to set standards.

Some attention to underrepresented groups occurs at the entry points. Campuses of the CSU, the largest preparer of teachers in the state, can admit up to 15 percent of their students who don’t meet the GPA requirement, and some LEAs reported giving a preference to underrepresented groups at the point of hiring. Once these teachers are admitted or employed, however, they must meet the same standards as other teachers. The greatest barrier to increasing the number of underrepresented groups in teaching is the basic skills proficiencies requirements.

Consistency

There are some criteria which are commonly applied across organizations at the same assessment point. IHEs are guided by the CTC program quality standards which specify skills to be demonstrated by program graduates. LEA criteria for teacher evaluation must include those specified in the Stull Act. However, interpretation of the criteria and standards applied vary both within and across institutions. Other common criteria used include liking children or youth (for IHEs) and enthusiasm (for LEAs).

The teacher descriptions of their assessments revealed a lack of consistency in the frequency and timing of assessments unrelated to the degree of problems experienced, both during student teaching and in the first year of teaching. The frequency of assessments by university supervisors ranged from two observations in 18 weeks to one per week. Some teachers reported daily feedback from their

master teachers, while others reported receiving almost no feedback at all. The timing of the first assessment made in the first year of teaching ranged from the first month of the school year to some time after January. The duration of observation by assessors ranged from 15 minutes to one hour.

Helpfulness

Helpfulness is directly related to the extent of the formative feedback provided by an assessment. The beginning teachers in the sample most commonly reported student teaching assessments to be helpful; however, approximately 25 percent of the teachers saw no value in their university supervisor's assessment and about 20 percent felt the same way about the assessment by the master teacher. The feedback of master teachers is generally valued over that of university supervisors because of the greater frequency of observation and their greater familiarity with the classroom and students. The sample teachers also reported that master teachers gave more practical, specific, and frequent suggestions, compared to university supervisors, and not only described, but modeled, alternative techniques and behaviors. General pedagogy was the area in which most teachers reported receiving useful feedback. First-year assessments were generally perceived by the teachers to be helpful because they provided a feeling of support and encouragement and an increase in confidence, rather than because they provided specific suggestions to improve teaching. A slight majority of teachers reported that the final evaluation made during their first year of teaching was helpful.

Communication of problems seems difficult for assessors. Many spoke of the need to preserve the self-esteem of the prospective or beginning teacher, who faces a considerable workload and stress. However, the result is that the negative feedback that teachers currently get, both in student teaching and in the first year of teaching, is often couched in the form of suggestions (*e.g.*, "Have you thought of trying . . .?") which do not necessarily identify weaknesses. Some teachers seemed to be unaware of their weaknesses because they are not specifically labeled as such.

In the context of a shortage of teachers and a profession where it is gradually being acknowledged that teaching skills can take years to develop, the best assessment system does little good to assess teachers unless they are supported in improving their skills. Every district in our sample provided some form of support for beginning teachers, with at least seven assigning beginning teachers to a specific support provider; in addition, 65 percent of the beginning teachers reported receiving guidance and assistance from an experienced teacher. However, the support providers in six of the seven districts clearly attempted to separate the evaluative role from support by observing only reluctantly in classrooms or observing only when requested to do so. Most support providers interviewed also reported frustration at a lack of time to support beginning teachers.

Although support providers tended to express concerns about the ability of beginning teachers to hear negative feedback, the beginning teachers interviewed

Teacher Assessment

expressed a desire for more feedback. They prefer that it be specific (*e.g.*, “Your initiation to the lesson captured the attention of almost all of your students.”) rather than general (*e.g.*, “You’re doing fine.”). They would also prefer to hear feedback from more than one person.

Conclusions

The assessment practices documented by the CNTP did not closely resemble an ideal assessment system. It is questionable whether the assessment practices conducted in a state with 54 county offices of education, 73 approved programs of professional preparation, and over 1,000 school districts could ever perfectly match the ideal. However, the CNTP findings noted that the uncoordinated patchwork of existing legislation regarding state assessment practices and the lack of a generally accepted framework for depicting effective teaching probably widened the disparity in assessment practices across institutions.

Rigorous evaluation of many aspects of teaching takes time and requires qualified, well-trained assessors. Both seem to be in short supply in many institutions. The past years have produced a steady round of budget cuts for IHEs and LEAs as California has experienced its worst recession since the Great Depression. Discretionary funding for expanding or enhancing IHE and LEA assessment practices has almost disappeared. Assessment models developed by FWL showed costs exceeding \$25 million dollars a year to assess 12,000 new teachers. These estimates did not include the costs of more rigorous pre-service assessments by IHEs.

Furthermore, the number of teachers to be assessed by IHEs and LEAs remains large. California public school populations continue to grow and the general population growth among those young enough to have children suggests that this growth will continue even if immigration of all kinds drops appreciably. When extremely large numbers of teachers must be assessed, the time available per teacher for assessment is significantly reduced, since supervision is labor intensive, and institutions seldom can increase their labor capacity proportionately. Assessment standards and the degree of assistance provided are also affected under these circumstances. The CNTP study revealed that a sort of “triage” system was practiced where the most needy got the most help, and those perceived to be doing okay were left alone to cope as best they could. Although all the respondents to the CNTP study acknowledged all new teachers need assistance, the system relied on teachers performing adequately to develop on their own, and was highly unlikely to push those teachers to enhance their teaching skills.

Although strapped for fiscal resources, California has mounted, through the BTSA program, a serious effort to address the concerns and recommendations that grew out of the CNTP studies. The BTSA program, currently a part of the regular funding base for public education in the state, provides for money to support beginning teachers and to undertake formative assessments, using innovative assessments drawn from the CNTP findings, as a means of increasing the knowl-

edge and skill development of these new teachers. While the funds are insufficient to provide this support and assessment to all new teachers in the state, currently there are 30 projects in place around the state, serving approximately 2,000 beginning teachers. As the economy of the state improves, the intent is to expand these programs to reach all new teachers in California with a two-year formal program of support and assessment.

California has also moved forward to develop a framework for defining the knowledge and skills that all teachers must possess to be effective in contemporary classrooms. It is being used (as a draft, not yet adopted) in all current BTSA projects to assist in creating a common language to guide new teacher development. State supported research groups are analyzing it in relation to other national teaching standards efforts such as the National Board for Professional Teaching Standards, the Interstate New Teacher Support and Assessment Consortium, and the recent revision of the standards of the National Council on the Accreditation of Teacher Education. Additionally, FWL is working to make the draft framework more feasible as a portfolio system for assessing and assisting classroom teachers.

Thus, under the auspices of SB 1422 (Bergeson, 1992), the CTC has been afforded the opportunity to investigate the further utilization of the findings of the CNTP in light of a redesign of basic teacher certification policies. The use of the framework to define better the minimum standards for completing a program of teacher preparation, the use of innovative assessments pilot tested in the CNTP as required elements within the program standards of teacher education programs, and the development of new modes of teacher certification that acknowledge formally the belief that becoming a teacher takes more than one year of coursework and student teaching, are all being “put on the table” for discussion by a broad array of organizations and individuals committed to improving the teaching profession in California.

Much remains to be done. The resource issues remain disturbingly powerful and there is great competition for limited funds. Serious methodological and policy issues must be investigated and thoroughly discussed before a thorough, comprehensive assessment system can be put into place. Nonetheless, the gradual movement toward a redefined path to teaching and assurances of quality envisioned by the Commons Commission in 1985 and continued in subsequent reform efforts has moved California toward a new view of teacher professionalism. This work, in turn, has enhanced the education of future generations of California children.

References

- Commission on Teacher Credentialing. (1994). *Teacher education, induction, and development for 21st Century schools: Preliminary plan for a review of teaching credential requirements*. Sacramento, CA: Author.
- California Commission on the Teaching Profession. (1985). *Who will teach our children? A strategy for improving California's schools*. Sacramento, CA: Author.

Teacher Assessment

Izu, J., Long, C., Stansbury, K., & Tierney, D. (1993) *Assessment components of the California New Teacher Project: Third year report*. San Francisco, CA: Far West Laboratory for Educational Research and Development.

Exploring Existing Teacher Assessment Practices in California: Implications for Potential Reforms

**By Jo Ann Izu, Claudia Long,
Kendyll Stansbury, & Dennis S. Tierney**

Jo Ann Izu, Claudia Long, Kendyll Stansbury, and Dennis S. Tierney were staff members at Far West Laboratory for Educational Research and Development, San Francisco, California, when the research described in this article was undertaken. Izu remains a staff member at Far West Laboratory. Long is currently an independent research consultant. Stansbury is now a staff member with the Connecticut Department of Education, Hartford, Connecticut. Tierney is now a consultant with the Professional Services Division of the California Commission on Teacher Credentialing, Sacramento, California.

Senate Bill 1422 (Bergeson, 1992) provides an unusually open-ended opportunity for all individuals interested in improving the preparation of classroom teachers in California. It calls upon the state's independent teaching standards board, the Commission on Teacher Credentialing (CTC), to "review the requirements for earning and renewing multiple and single subject (elementary and secondary) teaching credentials..." This statute allows significant latitude in defining the purposes, scope and issues to be studied as a part of the review. The CTC is interpreting this legislation as an opportunity to "invent a new system of teacher credentialing for 21st century schools" (CTC, November, 1994).

Teacher Assessment

This landmark legislation grew out of earlier efforts in California to address some long-standing concerns about teaching as a profession. Beginning with the California Commission on the Teaching Profession, known as the “Commons Commission Report,” in 1985, the state attempted to develop new policies to improve the teaching profession and thereby make the public schools better. This report, and others that followed, called for a restructuring of the teaching career and the establishment of more rigorous standards. In 1988, the Bergeson Act called for “new policies to govern the support and assessment of beginning teachers,” launching the California New Teacher Project (CNTP). The CNTP was a large scale, four-year pilot test of new approaches to supporting beginning teachers in the first two years of professional work, coupled with an evaluation of a wide array of teacher assessments intended to improve their practice through more detailed and meaningful feedback.

The Far West Laboratory for Educational Research and Development (FWL) was awarded a contract to conduct the research related to teacher assessments. This article reports the essential findings of one phase of the assessment studies conducted during the CNTP and suggests some implications from that study for the impending reform efforts embodied in the SB 1422 study. Additional research conducted under the auspices of the Beginning Teacher Support and Assessment (BTSA) project are summarized in accompanying articles in this special edition of *Teacher Education Quarterly* and in other publications.

The final phase of the FWL assessment studies examined the existing teacher assessment practices in California from the point of entry to a teacher education program through the awarding of tenure by a school district. Policymakers wanted to know more about the state of teacher assessment practices before launching new policy initiatives.

This article limits itself to the findings from that last phase of the CNTP research and relevant implications for future efforts at modifying the assessment practices across institutions and during the early years of a teacher’s professional development.

CNTP Assessment Studies

As part of the CNTP, co-sponsored by the CTC and the California Department of Education (CDE), FWL conducted a study of existing teacher assessment practices in California during the spring and early summer of 1991. The study focused on the evaluation of prospective and beginning teachers from undergraduate content area coursework through the awarding of tenure by a district.

The research design called for in-depth examinations of assessment practices in a small number of organizations which collectively represented the diverse contexts in California and at the same time maximized the diversity of assessment practices. Documents were examined and staff interviewed from 12 institutions of

higher education (IHEs), 12 districts (LEAs), and the CTC. Each different type of organization constituted a separate sub-study. In addition, 22 beginning teachers were interviewed about their experiences, both to provide an otherwise missing perspective and to see how the totality of assessment practices did and did not cohere. With the exception of the statewide examinations accepted or required by the CTC, each sub-study included teachers, campuses, and districts representing rural, suburban, and urban contexts, with size ranging from small to very large. The IHEs included campuses from the California State University (CSU) and University of California systems as well as private and independent colleges and universities. Some organizations were selected because they were known to have implemented innovative assessment practices. Beginning teachers represented a range of performance on various assessment instruments pilot tested in another phase of the CNTP.

In addition to examining relevant documents, the researchers interviewed at least one person in every role group that participated in key decision-making with respect to teacher evaluation. Administrators provided overviews of each assessment point for which they had responsibility. All respondents, including the beginning teachers, were asked to describe the processes in which they were directly involved. Twenty-one documents, on average, per IHE and up to 15 per LEA were examined. Interviews with 172 IHE staff and 152 LEA staff across the 24 organizations were conducted by trained interviewers.

Information was collected on eight points at which beginning teachers are often examined in their teaching career, including:

- 1) undergraduate coursework examinations in content areas;
- 2) admission to teacher preparation programs;
- 3) advancement to student teaching;
- 4) performance assessment in student teaching and other teacher education courses;
- 5) CBEST and other state-mandated examinations to satisfy credentialing requirements;
- 6) applications for employment;
- 7) formative assessment of new teachers for support purposes;
- 8) summative assessment of new teachers for continued employment decisions.

Site summary forms were completed for each IHE and LEA studied. The summaries highlighted both common and unusual policies and site-level practices. After writing summaries of findings for each separate sub-study, the four authors met and developed overall findings across sub-studies, which are described in the remainder of this article.

Focus of Assessments

Subject matter knowledge, which provides a foundation for the development of skills in content pedagogy, is the primary focus of the assessments for entry into

Teacher Assessment

a teacher preparation program. Assessment generally occurs through either a standardized examination approved by the state, or examinations in courses prescribed by a subject matter preparation program. Subject matter knowledge seldom gets re-examined at later points of assessment, as both IHEs and LEAs rely heavily on this initial screen. Subject matter knowledge may also be indirectly evaluated when IHEs and LEAs look at content pedagogy.

Current methods of assessing subject matter knowledge generally focus on breadth, not depth, of knowledge. They miss such deficiencies as inadequate knowledge of the structure of the discipline, an inability to sequence topics in a meaningful way, and a lack of ability to apply basic constructs (*e.g.*, fractions, different stages of the writing process). However, the revised subject matter tests recently adopted by the CTC include elements of application designed to address the last deficiency.

General pedagogy, especially classroom management, is the major focus of all assessments from student teaching through the early years of teaching. The state relies on IHE assessments of general pedagogy for credentialing requirements. Looking across the assessments conducted by the IHEs, LEAs, and the CTC, evaluation of skills in content pedagogy are consistently lacking in depth. During student teaching, content pedagogy is minimally assessed through observations and examination of lesson plans, and standards tend to be relatively low. The low standards do not seem to be grounded in a clearly defined perspective on teacher development. As with general pedagogy, the CTC again relies on IHEs to evaluate content pedagogy for credentialing purposes. Content pedagogy was also often minimally assessed by the LEAs in our sample at the point of hiring and/or in the first year of teaching. Skill in content pedagogy is most often defined by the LEAs as the ability to use the instructional techniques emphasized by the district.

Knowledge of students and of teaching diverse students is generally a very minor focus of all the assessments from student teaching on. Only a small amount of feedback during student teaching is focused on teaching diverse students, and tends to be directed toward classroom management issues or teaching students of differing ability levels. Knowledge of students is most directly assessed by LEAs at the point of hiring, when specific questions are sometimes asked about a teacher's familiarity or experience with certain types of students.

General academic ability and personal attributes are always assessed. General academic skills are evaluated at every entry point, usually through multiple measures: CBEST, GPA, and occasionally through performance assessments (*e.g.*, a writing sample). Personal attributes are primarily assessed through observation and interviews. Although IHEs and LEAs differ in some of the specific attributes they emphasize, almost all look for enthusiasm, rapport with children, and a commitment to teaching.

Forms of Assessments

Observation is the primary form of assessment used by IHEs and LEAs to assess a beginning teacher's knowledge, skills, and abilities during student teaching and the first years of teaching. Document review is commonly used at points of entry; however, IHEs and LEAs vary in how thoroughly they review materials and in the inferences they make from the same information. Written examinations are used by IHEs and the CTC to assess subject matter knowledge and basic skills proficiencies. IHEs also use written examinations to assess general and content knowledge, usually before the state of student teaching.

Alternative performance assessments (*e.g.*, unit/lesson plans, portfolios) are secondary forms of assessments sometimes used from teacher education coursework on. The CTCn has incorporated performance assessment into the subject matter knowledge examinations. Interviews are commonly conducted by IHEs at admission to the teacher preparation program and by LEAs at hiring. Alternative methods of assessment such as student achievement data and student evaluations of the teacher are occasionally used during student teaching and the first years of teaching.

Technical Quality

The technical quality of these assessment practices varies considerably both between and within institutions, but with the exception of the standardized state assessments, it is generally low. Clearly stated expectations and performance standards are rarely a characteristic of present assessments. Moreover, the assessments with the clearest criteria and standards are those that measure general academic ability and basic skills proficiencies; they do not measure teaching skills. Criteria used in present assessments of teaching skills are most often represented by broad categories whose interpretation is largely left up to individual assessors. Rating scales, when used, are rarely defined. In some small credential programs and districts, frequent conversations among the limited number of assessors increases the uniformity in interpretation and rating of criteria. In larger programs and districts, extensive training would be required to achieve the same result.

Not surprisingly, preparation of assessors, when it occurs, is largely oriented more to procedure than to substance. Even this extent of preparation varies by assessment point. For instance, training provided by districts for assessments for retention is common, while training for hiring assessments is not. Although beginning teachers interviewed said that most of the detailed and frequent feedback during student teaching is from cooperating (or master) teachers, the cooperating teachers themselves tend to receive no training for this role. Over one-third of the cooperating teachers interviewed reported some confusion over the meaning of some parts of the form they use to rate student teachers.

An Assessment System?

Desired characteristics of an ideal statewide assessment system are: rigor, thoroughness, consistency, and helpfulness. The extent to which these constructs characterize the assessment practices identified in the overall study is discussed separately for each construct.

Rigor

Both IHE and LEA assessment practices are uneven in terms of rigor, both across institutions and within institutions and credential programs. Evaluation criteria are identified and sometimes defined by detailed examples. However, standards, *i.e.*, the rating categories used, are rarely defined. Those IHEs and LEAs that define the constructs being assessed, either conceptually or by providing examples, usually have the most rigorous assessments as well.

Few documents examined reflected a well-articulated perspective on the development of teaching skills. The form and procedures for evaluation of teachers at an early stage (*e.g.*, student teacher in first placement, first-year teacher) were almost always the same as that for teachers at later stages (*e.g.*, student teacher in second placement, 20-year veteran). Assessors reported taking the extent of a teacher's experience into account in their evaluation, but the basis for doing so was not clear, and in many cases appeared to vary with the individual assessor.

LEAs differ in their capacity to maintain high standards in the evaluation of prospective and beginning teachers. Standards for new teacher assessment by LEAs are driven by market forces, where beginning teachers are judged relative to those who might be expected to replace them. Districts vary in their attractiveness to teachers, and some of the most influential negative factors (*e.g.*, inner city location, per-pupil funding levels) are beyond a district's control. Although recruitment strategies have some impact, a district's attractiveness strongly influences the number and quality of applicants. As a result, districts with many high quality applicants are able to apply high standards, while districts with difficulty recruiting credentialed teachers are not. However, districts with severe staffing problems do employ standards by which they reject credentialed teachers. These standards often include more emphasis on knowledge of and empathy with the type of students in the district and less emphasis on pedagogy, reflecting a belief that pedagogy is more easily taught.

In terms of identifying candidates who should be eliminated from the profession, the assessment is most rigorous at the entry points. From the perspective of the beginning teachers interviewed, the rigor of assessment depends on the variety of areas being assessed, the amount of work expected, and the level of expectations held by the assessor(s). Most considered their student teaching assessments to be more rigorous than those experienced during the first years of teaching.

The rigor of assessments was affected both by resources and available

methodology. State assessments exhibit high technical quality, in part because the high cost of well-developed assessments can be spread across a large number of candidates, and a source of income is available to fund administration, development, and implementation costs.

For eliminating candidates from the system, IHEs and LEAs tend to favor assessments that are “appeal-proof,” which tends to mean one of two characteristics. The first characteristic is that the assessment criteria are directly measurable. For example, while the validity of using a GPA as a measure of academic skills may be argued, whether or not a specific GPA meets the standard cannot. However, whether or not a teacher is using an appropriate instructional technique is less clear, and more vulnerable to conflicting interpretations. The ease with which indicators of skills can be measured varies across teaching domains. Classroom management is relatively easy to observe and document, while content pedagogy and knowledge of students are not.

The second, and related, characteristic of an “appeal-proof” assessment is the degree to which there is a consensus on a definition of important teaching skills across contexts and teaching styles. One reason that content pedagogy and knowledge of students are difficult to document is that they are difficult to define across different teaching contexts and styles.

LEAs are not required to give reasons for releasing teachers prior to receipt of tenure, and assessors in many of the sample districts stated that they did not explain decisions not to re-hire beginning teachers to avoid potential legal problems. Some districts even have recommended guidelines based on legal criteria as to what assessors should and should not tell teachers at the point of dismissal.

Thoroughness

Assessments seem to be conducted without the benefit of a general teaching framework which serves as a context in which to discuss a beginning teacher’s knowledge, skills, and abilities. Often there are lists of competencies, but is not clear how the competencies interrelate. Some items seem to be more characteristic of experienced than student or beginning teachers (*e.g.*, for a student teacher, “paces individual lessons appropriately”). Without a general teaching framework, each teacher is left to generalize from their particular teaching experiences and invent their own framework. Prospective and beginning teachers seem to have difficulty in putting feedback about specific actions and reactions into a larger framework so that they see alternative ways of approaching problems. The lack of a framework also facilitates the rejection of negative feedback, because the teacher is unable to link that specific feedback to a larger principle or issue. Thus, negative feedback is more vulnerable to being ignored, or even rejected due to being attributed to a personality conflict or lack of contextual understanding by the assessor.

Information from previous assessments is rarely used for diagnostic purposes, both across institutions (*e.g.*, from IHE to LEA) and within institutions (*e.g.*,

Teacher Assessment

information from education coursework used in student teaching; formative and summative evaluations of new teachers). Where information from the IHE is available to the LEA for hiring of new staff, a lack of time to consider the information means that it most often either is not used or is used only in the final stages of the selection process, where the number of active candidates is considerably reduced. Even if assessors or institutions are willing to share and use information, there are legal liabilities with respect to negative information which affects the information sent. Generally, only “fatal” weaknesses of applicants are disclosed. Minor to major weaknesses are often identified by omission, requiring an evaluator to go well beyond what is written in letters of recommendation. Typical comments were: “Unless the letter of recommendation specifically states that classroom management is strong, we assume that it isn’t,” and “Listen to what they don’t say as much as what they do say.”

Fairness

The absence of a conceptual framework of teaching makes it difficult to set fair levels of expectations. Because of the lack of explication of assessment standards, fairness often depends on the professional judgment of people who are not always well trained. The differences in teaching contexts, especially for IHEs, also contribute to the difficulty of maintaining fairness, perhaps contributing to the reluctance to set standards.

Some attention to underrepresented groups occurs at the entry points. Campuses of the CSU, the largest preparer of teachers in the state, can admit up to 15 percent of their students who don’t meet the GPA requirement, and some LEAs reported giving a preference to underrepresented groups at the point of hiring. Once these teachers are admitted or employed, however, they must meet the same standards as other teachers. The greatest barrier to increasing the number of underrepresented groups in teaching is the basic skills proficiencies requirements.

Consistency

There are some criteria which are commonly applied across organizations at the same assessment point. IHEs are guided by the CTC program quality standards which specify skills to be demonstrated by program graduates. LEA criteria for teacher evaluation must include those specified in the Stull Act. However, interpretation of the criteria and standards applied vary both within and across institutions. Other common criteria used include liking children or youth (for IHEs) and enthusiasm (for LEAs).

The teacher descriptions of their assessments revealed a lack of consistency in the frequency and timing of assessments unrelated to the degree of problems experienced, both during student teaching and in the first year of teaching. The frequency of assessments by university supervisors ranged from two observations in 18 weeks to one per week. Some teachers reported daily feedback from their

master teachers, while others reported receiving almost no feedback at all. The timing of the first assessment made in the first year of teaching ranged from the first month of the school year to some time after January. The duration of observation by assessors ranged from 15 minutes to one hour.

Helpfulness

Helpfulness is directly related to the extent of the formative feedback provided by an assessment. The beginning teachers in the sample most commonly reported student teaching assessments to be helpful; however, approximately 25 percent of the teachers saw no value in their university supervisor's assessment and about 20 percent felt the same way about the assessment by the master teacher. The feedback of master teachers is generally valued over that of university supervisors because of the greater frequency of observation and their greater familiarity with the classroom and students. The sample teachers also reported that master teachers gave more practical, specific, and frequent suggestions, compared to university supervisors, and not only described, but modeled, alternative techniques and behaviors. General pedagogy was the area in which most teachers reported receiving useful feedback. First-year assessments were generally perceived by the teachers to be helpful because they provided a feeling of support and encouragement and an increase in confidence, rather than because they provided specific suggestions to improve teaching. A slight majority of teachers reported that the final evaluation made during their first year of teaching was helpful.

Communication of problems seems difficult for assessors. Many spoke of the need to preserve the self-esteem of the prospective or beginning teacher, who faces a considerable workload and stress. However, the result is that the negative feedback that teachers currently get, both in student teaching and in the first year of teaching, is often couched in the form of suggestions (*e.g.*, "Have you thought of trying . . .?") which do not necessarily identify weaknesses. Some teachers seemed to be unaware of their weaknesses because they are not specifically labeled as such.

In the context of a shortage of teachers and a profession where it is gradually being acknowledged that teaching skills can take years to develop, the best assessment system does little good to assess teachers unless they are supported in improving their skills. Every district in our sample provided some form of support for beginning teachers, with at least seven assigning beginning teachers to a specific support provider; in addition, 65 percent of the beginning teachers reported receiving guidance and assistance from an experienced teacher. However, the support providers in six of the seven districts clearly attempted to separate the evaluative role from support by observing only reluctantly in classrooms or observing only when requested to do so. Most support providers interviewed also reported frustration at a lack of time to support beginning teachers.

Although support providers tended to express concerns about the ability of beginning teachers to hear negative feedback, the beginning teachers interviewed

Teacher Assessment

expressed a desire for more feedback. They prefer that it be specific (*e.g.*, “Your initiation to the lesson captured the attention of almost all of your students.”) rather than general (*e.g.*, “You’re doing fine.”). They would also prefer to hear feedback from more than one person.

Conclusions

The assessment practices documented by the CNTP did not closely resemble an ideal assessment system. It is questionable whether the assessment practices conducted in a state with 54 county offices of education, 73 approved programs of professional preparation, and over 1,000 school districts could ever perfectly match the ideal. However, the CNTP findings noted that the uncoordinated patchwork of existing legislation regarding state assessment practices and the lack of a generally accepted framework for depicting effective teaching probably widened the disparity in assessment practices across institutions.

Rigorous evaluation of many aspects of teaching takes time and requires qualified, well-trained assessors. Both seem to be in short supply in many institutions. The past years have produced a steady round of budget cuts for IHEs and LEAs as California has experienced its worst recession since the Great Depression. Discretionary funding for expanding or enhancing IHE and LEA assessment practices has almost disappeared. Assessment models developed by FWL showed costs exceeding \$25 million dollars a year to assess 12,000 new teachers. These estimates did not include the costs of more rigorous pre-service assessments by IHEs.

Furthermore, the number of teachers to be assessed by IHEs and LEAs remains large. California public school populations continue to grow and the general population growth among those young enough to have children suggests that this growth will continue even if immigration of all kinds drops appreciably. When extremely large numbers of teachers must be assessed, the time available per teacher for assessment is significantly reduced, since supervision is labor intensive, and institutions seldom can increase their labor capacity proportionately. Assessment standards and the degree of assistance provided are also affected under these circumstances. The CNTP study revealed that a sort of “triage” system was practiced where the most needy got the most help, and those perceived to be doing okay were left alone to cope as best they could. Although all the respondents to the CNTP study acknowledged all new teachers need assistance, the system relied on teachers performing adequately to develop on their own, and was highly unlikely to push those teachers to enhance their teaching skills.

Although strapped for fiscal resources, California has mounted, through the BTSA program, a serious effort to address the concerns and recommendations that grew out of the CNTP studies. The BTSA program, currently a part of the regular funding base for public education in the state, provides for money to support beginning teachers and to undertake formative assessments, using innovative assessments drawn from the CNTP findings, as a means of increasing the knowl-

edge and skill development of these new teachers. While the funds are insufficient to provide this support and assessment to all new teachers in the state, currently there are 30 projects in place around the state, serving approximately 2,000 beginning teachers. As the economy of the state improves, the intent is to expand these programs to reach all new teachers in California with a two-year formal program of support and assessment.

California has also moved forward to develop a framework for defining the knowledge and skills that all teachers must possess to be effective in contemporary classrooms. It is being used (as a draft, not yet adopted) in all current BTSA projects to assist in creating a common language to guide new teacher development. State supported research groups are analyzing it in relation to other national teaching standards efforts such as the National Board for Professional Teaching Standards, the Interstate New Teacher Support and Assessment Consortium, and the recent revision of the standards of the National Council on the Accreditation of Teacher Education. Additionally, FWL is working to make the draft framework more feasible as a portfolio system for assessing and assisting classroom teachers.

Thus, under the auspices of SB 1422 (Bergeson, 1992), the CTC has been afforded the opportunity to investigate the further utilization of the findings of the CNTP in light of a redesign of basic teacher certification policies. The use of the framework to define better the minimum standards for completing a program of teacher preparation, the use of innovative assessments pilot tested in the CNTP as required elements within the program standards of teacher education programs, and the development of new modes of teacher certification that acknowledge formally the belief that becoming a teacher takes more than one year of coursework and student teaching, are all being "put on the table" for discussion by a broad array of organizations and individuals committed to improving the teaching profession in California.

Much remains to be done. The resource issues remain disturbingly powerful and there is great competition for limited funds. Serious methodological and policy issues must be investigated and thoroughly discussed before a thorough, comprehensive assessment system can be put into place. Nonetheless, the gradual movement toward a redefined path to teaching and assurances of quality envisioned by the Commons Commission in 1985 and continued in subsequent reform efforts has moved California toward a new view of teacher professionalism. This work, in turn, has enhanced the education of future generations of California children.

References

- Commission on Teacher Credentialing. (1994). *Teacher education, induction, and development for 21st Century schools: Preliminary plan for a review of teaching credential requirements*. Sacramento, CA: Author.
- California Commission on the Teaching Profession. (1985). *Who will teach our children? A strategy for improving California's schools*. Sacramento, CA: Author.

Teacher Assessment

Izu, J., Long, C., Stansbury, K., & Tierney, D. (1993) *Assessment components of the California New Teacher Project: Third year report*. San Francisco, CA: Far West Laboratory for Educational Research and Development.