Authentic assessment of teaching in context

Linda Darling-Hammond[a],*, Jon Snyder[b]

[a]*Stanford University School of Education, CERAS 402, 520 Galvez Mall, Stanford, CA 94305, USA*
[b]*University of California, Santa Barbara, Santa Barbara, CA, USA*

Abstract

The demands of teaching more challenging content to more diverse learners suggest a need for teacher education that enables teachers to become more sophisticated in their understanding of the effects of context and learner variability on teaching and learning. Instead of implementing set routines, teachers need to become ever more skillful in their ability to evaluate teaching situations and develop teaching responses that can be effective under different circumstances. This article examines how a growing number of teacher education programs are using authentic assessments of teaching – cases, exhibitions, portfolios, and problem-based inquiries (or action research) – as tools to support teacher learning for these new challenges of practice. Using specific teacher education programs as examples, the article examines how and why these strategies appear to provide support for teacher learning and avenues for more valid assessment of teaching. The authors also discuss circumstances in which these strategies may be less effective and suggest features of the assessments and programmatic contexts that are associated with more and less successful use. © 2000 Elsevier Science Ltd. All rights reserved.

Keywords: Teacher assessment; Teacher education; Authentic assessment

Practitioners and researchers engaged in education reform suggest that teaching is becoming more complex in response to increasingly challenging curriculum expectations and growing diversity among students (Darling-Hammond, 1997; Shulman, 1987; Sizer, 1992). The process of teaching for understanding and application rather than for rote recall creates greater unpredictability in teaching as teachers must be able to understand and capitalize upon student thinking in order to manage a process of knowledge construction that is different for each one. And as schools include

a greater range of students from different backgrounds and with different approaches to learning, formulas for teaching that do not take account of students' experiences and needs are less and less successful (Darling-Hammond, 1997). Teachers who want to teach diverse learners effectively must address the distinctive resources each brings to the table, including different dispositions, prior experiences and knowledge, cultural and linguistic capital, and sources of potential identification and opposition (Doyle, 1979; Hollins, 1989; Moll, 1988).

The desire to succeed at much more ambitious learning goals with a much more diverse student population also creates new challenges for teacher preparation. If all students pursued an identical path to understanding, learning might be ensured by curricular design alone. Teachers could be

*Corresponding author. Tel.: + 1 650 723-3555; fax: + 1 650 723-7578.

E-mail address: LDH1222@aol.con (L. Darling-Hammond).

prepared to implement curriculum using a reasonably standard range of teaching techniques. However, teaching that responds to human diversity and aims for cognitive flexibility requires a wide range of teaching strategies that are activated by sophisticated judgments grounded in disciplined experimentation, insightful interpretation of (often ambiguous) events, and continuous reflection. This kind of teaching aims to diagnose and make use of variability, rather than implement uniform techniques or routines.

This paper examines the growing use of teacher education curriculum and assessment strategies intended to prepare teachers for these challenges. An expanding number of teacher education programs are using authentic assessments of teaching as one set of tools to help novice teachers create, in a principled fashion, bridges from generalizations about practice to apparently idiosyncratic, contextualized instances of learning. Under the title of authentic assessment, we include opportunities for developing and examining teachers' thinking and actions in situations that are experience based and problem oriented and that include or simulate actual acts of teaching. Such acts of teaching include plans for and reflections on teaching and learning, as well as activities featuring direct interaction with students.

Our reference to *contextualized* teaching and learning is meant to underscore the fact that all teaching and all learning is shaped by the contexts in which they occur. These contexts are defined by the nature of the subject matter, the goals of instruction, the individual proclivities and understandings of learners and teachers, and the settings within which teaching and learning take place. Such variables as school organization, resources, materials, amount of time and how it is structured for learning, the duration and nature of relationships among students and teachers, community norms and values influence the processes and outcomes of teaching decisions. The extent to which context influences teaching – and determines what kinds of approaches to teaching will be effective – is a factor that is increasingly acknowledged in research on teaching, in teacher education, and in the assessment of teaching.

Many teacher education programs, along with organizations like the National Board for Profes-

sional Teaching Standards (NBPTS), have been engaged over the past decade in developing assessments of teaching and learning that take these issues of context into account. Cases, portfolios that assemble artifacts of practice, exhibitions of performance, and problem-based inquiries are among the tools being used to try to capture important attributes of teaching and reasoning about teaching. These tools allow the application of theoretical principles to problems in specific contexts while appropriately complicating efforts to draw generalizations about practice. In doing so, they may also transform teachers' understandings of both theory and practice.

A small but growing body of research suggests that such strategies can help teachers understand more deeply the many variables that influence their work. In the case of cases and portfolios that require teachers to examine student learning in relation to their teaching, for example, teachers claim that the process of engaging in such analysis ultimately enriches their ability to understand the effects of their actions and helps them better meet the needs of diverse students (Athanases, 1994; Bliss & Mazur, 1997; Bradley, 1994; Darling-Hammond, 1999; Haynes, 1995; Ingvarson & Marrett, 1997; Tracz et al., 1995). On the other hand, such assessments may achieve these outcomes only in certain circumstances (for example, when they are designed in particular ways and/or used by teachers who have already achieved certain kinds of skills or understandings), and they may have other limitations requiring attention from those who would use them.

What do we know about cases, exhibitions, portfolios, and inquiries that appear to be powerful in promoting teacher learning? What features of these strategies may contribute to the development of teachers' professional judgement and to the more valid assessment of teaching? Are there features of such assessments that can undermine or enhance their potential for supporting or examining teacher learning? In this paper, we describe how these authentic assessment tools are used in different teacher education programs to address some of the fundamental dilemmas of learning to teach in a world where many factors influence learning. We bring to bear evidence from a collection of recent case

studies of extraordinary teacher education programs conducted by the National Commission on Teaching and America's Future, as well as evidence about other programs recently reported in the literature. Beyond our descriptions of these approaches – each of which is widely used in many other teacher education programs – we examine how the nature of the assessment strategy appears to shape and reflect the learning of prospective teachers and their capabilities as teachers.

1. A rationale for contextualized assessment of teaching

A major problem of teaching and teacher education is the problem of moving from intellectual understanding to enactment in practice (Kennedy, 1999). The problem of enactment, especially in light of current expectations for teaching, is not trivial. As Villegas (1997) notes:

> Because teaching must build upon and modify students' prior knowledge, responsive teachers select and use instructional materials that are relevant to students' experiences outside school (Hollins, 1989), design instructional activities that engage students in personally and culturally appropriate ways (Garibaldi, 1992; Irvine, 1990), make use of pertinent examples or analogies drawn from the students' daily lives to introduce or clarify new concepts (Irvine, 1992), manage the classroom in ways that take into consideration differences in interaction styles (Tikunoff, 1985), and use a variety of evaluation strategies that maximize students' opportunities to display what they actually know in ways that are familiar to them (Moll, 1988; Ortiz & Maldonado-Colon, 1986) (p. 265).

Teaching in ways that are responsive to students requires that teachers be able to engage in systematic learning from teaching contexts as well as from more generalized theory about teaching and learning. Without an understanding of how culture, experience, readiness, and context influence how people grow, learn, and develop, it is difficult for teachers to make good judgements about how to deal with the specific events in the classroom. However, without an appreciation for the intense, interactive realities of classroom life, and for the multidimensional problems and possibilities posed by individual learners, it is difficult for the theoretically knowledgeable to apply what they know in practice. It is both more difficult to develop such abilities and to evaluate them than it is to assume a single approach to teaching or a single right answer to teaching problems. Those who have worked to develop assessments of such complex performances argue that traditional paper and pencil measures and low-inference observation tools are insufficiently context sensitive to assess teaching that is effective for diverse learners (Darling-Hammond, Wise & Klein, 1998; Haertel, 1990; Shulman, 1987).

Studies of the predictive validity of traditional multiple-choice paper and pencil tests of teaching (for example, the National Teacher Examinations) have found little evidence that such tests are correlated with teacher ratings or teachers' classroom effectiveness (Andrews, Blackmon & Mackey, 1980; Ayers & Qualls, 1979; Haney, Madaus & Kreitzer, 1987; Quirk, Witten & Weinburg, 1973.) One reason for the lack of predictive validity may be that the ability to recognize information when it is presented in a list of responses is significantly different from the ability to produce the same kind of analysis or to enact corresponding ideas in practice. Another reason may be that the tests in use have featured decontextualized teaching scenarios (Darling-Hammond, 1986) and have not represented core tasks of teaching in ways that accurately reflect their conduct in classrooms, including the integration of multiple strands of knowledge and skill (Haertel, 1990; Shulman, 1987).

Systems of teacher observation in classrooms are, in one way, potentially more authentic. They do, in fact, look at teaching as it is being enacted. However, the generation of low-inference rating systems developed in the 1980s presumed a single short list of observable teaching behaviors that could be deemed "effective" (e.g. keeps a brisk pace of instruction, manages routines, develops behavioral objectives) and could be tallied from countable indicators. These approaches did not address important differences in context and content, and

they ignored the influence of teaching on learning. Because effective teachers vary their behaviors across teaching situations, evaluation results based on low-inference behavioral instruments have low generalizability (Shavelson & Dempsey-Atwood, 1976; Stodolsky, 1984). Furthermore, the lists of teaching behaviors proved to have questionable relationship to other measures of teaching effectiveness. Context-related differences in the outcomes associated with a given list of behaviors produced inconsistent findings in process-product studies and undermined confidence in simple translations of results into teacher evaluation instruments (Darling-Hammond et al., 1998).

Some studies have found that teachers who learn to teach to decontextualized evaluation tools constructed around a set list of teaching behaviors consider a narrower range of teaching concerns (Hoover & O'Shea, 1987). In addition, they are less likely to attend to issues of curriculum planning, of content pedagogy, of the relationship between their practices and student responses or outcomes, or to teaching tasks that fall outside of the observation context (Darling-Hammond & Sclan, 1992; French, Hodzkom & Kuligowski, 1990). As Floden and Klinzing (1990) note: "Training teachers to follow a fixed set of prescriptions discourages teachers from adapting their instruction to the particular subjects and students they are teaching. Hence, the instructional effectiveness of teachers given such training is unlikely to be at a high level" (pp. 16, 17).

There is growing interest among educators and evaluators in constructing other forms of assessment that better reflect the complexity of teaching and can provide valid data about competence while helping teachers improve the caliber of their work with children and their families. In the rest of this article we describe several approaches to the authentic assessment of teaching and describe how these practices are currently being used in pre-service teacher education programs in the United States. We provide data on the perceived effects of these tools on candidates and programs, and we conclude the paper with a discussion of key practice, research, and policy issues arising from the use of authentic assessment in teacher education programs.

The group of programs described here is neither random nor representative, but illustrative of institutions that have relatively highly developed practices in this area. It derives from two sources. One source of information is a study of seven extraordinarily effective teacher education programs conducted for the National Commission on Teaching and America's Future (Darling-Hammond, 2000).[1] These programs were selected through reputational sampling from scholars and practitioners in the field, surveys of graduates and employers (in comparison to a random sample of beginning teachers in the United States), and observations of graduates. The programs show evidence of regularly producing teachers who are unusually well-prepared to work effectively with diverse learners. The study found that among the practices the programs have in common is the extensive use of cases, portfolios, exhibitions, and action research inquiries as tools for developing and assessing teaching skills. A second source of program information is the literature on assessment in teacher education which provides evidence about practices and potential outcomes of some well-described examples. This literature also suggests that the practices described here are becoming increasingly commonplace in teacher education programs, although strategies for using them certainly vary. As a consequence, it seems timely to consider how these approaches operate, why programs find them useful, and what features may be important to their success in developing and assessing context-sensitive teaching.

2. A framework for defining authentic assessment of teaching

Below we outline four aspects of authentic assessment of teaching that appear from emerging research to be important both for measuring teaching and enhancing candidates' abilities to teach well.

---

[1] The programs are at Alverno College, Milwaukee, WI; Bank Street College, New York, NY; Trinty University, San Antonio, TX; University of California, Berkeley; University of Southern Maine; University of Virginia; and Wheelock College, Boston, MA.

1. *Assessments sample the actual knowledge, skills, and dispositions desired of teachers as they are used in teaching and learning contexts, rather than relying on more remote proxies*. Although some contexts for assessment may be a step removed from daily classroom life, the tasks undertaken require the integration and use of knowledge and skills as they are employed in practice. Assessment tasks include actual examples of the work of teaching (videotapes of teaching, plans, and assessments of student learning, for example) and analyses of teaching, learning, and curriculum or materials. Such assessments seek, to varying degrees, to deal with the problem of enactment; that is, the fact that talking or writing about teaching, or recognizing answers to questions about teaching, cannot fully predict a person's capacity to succeed in the complex realities of actual teaching in the classroom. This criterion reflects the common sense notion that if one wants to assess a performance skill like swimming, for example, it is useful to have the swimmer in the water at some point.

It is, perhaps, worth re-emphasizing that the authentic assessment of teaching does not consist entirely of classroom observation. Many aspects of teaching are only indirectly visible during the classroom portion of a teacher's practice. These include planning regarding how to represent content and how to adapt lessons to the needs of particular learners, work with children and their families and the community outside of classroom hours, analyses of the special strengths and challenges students possess, and work with colleagues on the planning and integration of instruction and on student and school-level problem solving, to name but a few. In these cases, classroom observations are actually remote proxies for the actual knowledge, skills, and dispositions to be assessed. Therefore, assessment tools such as interviews, teacher reflections and analyses, samples of feedback the teacher provided students and/or that others provided the teacher, and other artifacts that represent these aspects of practice may better meet the underlying principle of authentic assessment.

2. *Assessments require the integration of multiple kinds of knowledge and skill as they are used in practice*. One complaint about traditional teacher education has been that students experience frag-

mentation among courses that treat different subject matter (content and pedagogy, for example, or learning, curriculum, and assessment, for another) as well as fragmentation in occasions for dealing with theory and practice (see, e.g. Goodlad, 1990). This leaves much of the task of assembling a knowledge base for teaching to the student, with resulting gaps and problems of translation. Proponents of new approaches argue that assessments that mirror teaching by seeking to integrate areas of knowledge used in combination can help forge these connections while better representing the tasks teachers must actually perform (Darling-Hammond et al., 1998).

In addition, use of such assessments may shape professional preparation programs in ways that encourage better integration of knowledge within and across courses and other learning experiences. So, for example, an assessment of a child's literacy development might rely upon study of research and theory about literacy development, learning, curriculum, and assessment; instruction in the use of literacy assessment tools and instructional strategies; practice and coaching in the collection and analysis of data about children's literacy learning; and reflection upon the data collected, its meaning, and implications for instruction. Incorporating such assessments into the ongoing curriculum of teacher preparation may heighten the probability that knowledge and skills will be better integrated and applied and that complex assessments will become practically feasible. If such assessments are treated largely as add-ons at the end of a course or program rather than as integral components of ongoing curriculum and instruction, the time, labor, and expense of conducting them could become overwhelming within the institutional constraints of teacher education programs. This condition also helps to ensure that the necessary opportunities for learning are present, thus enhancing the probabilities of success.

3. *Multiple sources of evidence are collected over time and in diverse contexts*. There are two principles underlying this criterion. The first is that assessments that support sound decisions require evidence based on adequate samples of thinking and behavior. An isolated sample of performance or a single genre of data is insufficient to inform

judgements about learning, teaching, program development, or candidate competence. Robust assessments of knowledge and skill, like those of the National Board for Professional Teaching Standards, can include written analyses, observation data (e.g., from a supervisor's, cooperating teacher's, or principal's observation), and performance samples such as videotapes, samples of student work from the student teacher's classroom, and samples of communications with families. Not only is the multiplicity of data sources important but the ways in which they are assembled is also critical. To understand a classroom event, one needs relevant information about students and their prior learning and approaches as well as about the event itself and about the teacher's decision-making processes – i.e. what goals she is trying to achieve and what aspects of content, student needs, and classroom or community context she is taking into account. In addition, since context does matter, assessments should provide candidates with opportunities to show their abilities in different settings, with different students, and with different lesson content. If teacher education is professional education, it should prepare candidates to take into account the different needs of students and demands of subject matter and other context variables when they are making decisions. The conscious effort of differentiating and analyzing the factors represented in different settings for practice is what distinguishes preparation for professional practice from an apprenticeship model in which novices aim to copy the skills of a veteran practitioner, as though they will be applicable in all contexts. The literacy assessment described above could meet this criterion, if candidates completed it a second time in a different context with a child of different needs and abilities. Explicitly recognizing the importance of context differences in assessments could help candidates develop more finely tuned perceptual and analytic abilities, reinforce the development of more professionally oriented and context-sensitive preparation programs, and advance the representation of different sociocultural and other contexts in performance assessments.

4. *Assessment evidence is evaluated by individuals with relevant expertise against criteria that matter for performance in the field*. This criterion acknow-

ledges that the bases for making inferences from an assessment are as important to its validity as is the assessment task itself. Two people looking at the same evidence base might draw entirely different conclusions about its meaning if they have different levels or kinds of expertise or if they are applying different expectations for what constitutes a good or competent performance. Newly developed performance assessments like those used by the National Board for Professional Teaching Standards or the Interstate New Teacher Assessment and Support Consortium (INTASC) are developed around standards for accomplished practice that reflect current research in the field and are scored by individuals with demonstrated expertise in the subject matter area and level of the candidate being assessed. A defensible basis for creating expert consensus about the meaning of teaching events or strategies is an important component of a performance assessment that can be used to evaluate or to guide the development of teaching skills.

Within the context of teacher education, one might posit a fifth requirement for authentic assessments that aim to develop as well as measure teaching judgement and skill: that the assessment practice includes multiple opportunities for learning and practicing the desired outcomes and for feedback and reflection. This criterion reflects a new expectation for the purposes of assessment: that is, that it helps develop competence, not just measure outcomes (Darling-Hammond et al., 1998). If one believes that a teacher is not something one becomes but rather something one is constantly becoming, then a core function of teacher education is to increase the ability of candidates to reflect upon and learn *from* teaching. Assessments of the work of a teacher that include opportunities for learning from feedback and reflection both support the development of greater levels of competence and measure a critical attribute of an effective teacher; the ability to learn from practice.

3. Tools for authentic assessment

We examine four tools that, in certain circumstances, meet the conditions outlined above: cases,

exhibitions of performance, portfolios, and problem-based inquiries (also known as action research). Each provides a means for organizing curriculum and instruction in a teacher education program as well as for assessing prospective teachers' developing abilities. No single one of these tools represents the totality of teaching. Each, however, assesses important aspects of teaching and, in interesting ways, reflects a different metaphor for teaching. Cases, for example, develop and assess teachers' abilities as decision makers. Exhibitions draw upon the performances of teaching and reflect the teacher as an artist. Portfolios support the teacher as a continuous learner who reflects on practice. Research and inquiry develop teachers as social science analysts. Used in combination, as many teacher education programs employ them, such tools allow novices to integrate different areas of learning and to apply them in different ways that, together, include many dimensions of a professional teaching role.

A counter image is that of the beginning teacher as a floundering navigator, someone whose entry into teaching is a sink or swim affair. Research on teacher development (Fuller, 1969; Katz, 1972) has suggested that beginning teachers may not be capable of advanced practice until they successfully move through concerns related to rudimentary survival and classroom management. However, the experience of highly successful programs using these highly scaffolded tools for developing and assessing teaching suggests that beginning professionals are capable of much more sophisticated practice than previously thought (Darling-Hammond, 2000; Darling-Hammond & Macdonald, 1999; Koppich, 1999; Lippincott, 1999; Merseth & Koppich, 1999; Miller & Silvernail, 1999; Snyder, 1999; Whitford, Ruscoe & Fickel, 1999; Zeichner, 1999). The caliber of the work generated by teacher candidates in programs where study is both rooted in practice and unremittingly analytic suggests that the concerns of beginning professionals can move more quickly from a focus on self to a focus on students when they have tools to help them train their sights on the effects of their actions and decisions.

3.1. Cases

In the preface to their book on *The Case for Education*, Colberg, Trimble and Desberg (1996) note that the growing interest in using case methods in teacher education can be explained with one word: "Context". Cases add context to theory. Whether they take the form of case reports – first-person narratives of personal experiences of teaching – or case studies – third person analyses of situations or students – cases allow the exploration of precepts, principles, theories, and perennial issues as they actually occur in the real world. Students may read and analyze cases, seeking the lessons and insights they offer, or they may write their own cases, developing interpretations of events as they work through the process of representing their experience. These efforts can motivate learning, serve as instructional material for others, and provide "antidotes to the dangers of overgeneralization" (Shulman, 1992, p. 3). Typically, cases represent instances of teaching and learning that pose dilemmas, provide carefully assembled evidence or data, and, sometimes, describe the outcomes of various decisions in specific situations. Contexts for cases may be defined by the nature of the subject matter and students; the history of a class, an event, or an individual; and the situations observed or strategies attempted.

Some cases provide visions of the possible – compelling sagas that can inspire and guide. Other cases describe the collision between design and chance and the surprises that are the essence of teaching experience. Cases may be developed from any number of perspectives. For instance, they may start from a subject matter perspective, probing teachers' understandings of curriculum and instruction by examining how teachers configure and analyze learning experiences aimed at the mastery of certain skills and content in light of student needs and classroom conditions. They may also arise from a student perspective, assessing teachers' knowledge and their skills of observation and interpretation by examining how teachers evaluate student learning and development in terms of strengths, interests, and needs. They may start from a cultural perspective, allowing teachers to inquire into students' lives and contexts in order to prepare

teachers for the intellectually and emotionally demanding experiences that can arise in culturally diverse classrooms and communities.

Proponents argue that, through careful analysis of cases and the inevitable puzzles that must be addressed between aspiration and accomplishment, teacher candidates develop theories and strategies. Skeptics may be concerned that highly contextualized accounts of seemingly idiosyncratic situations may fail to add up to more principled understanding of teaching concerns, or that the use of cases may add only to the lore of teaching decisions based on personal opinions uninformed by broader professionwide knowledge. This legitimate concern suggests that the utility of cases as tools for teaching or assessment may depend substantially on what kinds of readings, discussions, analyses, and commentaries are wrapped around them as they are constructed or evaluated.

3.2. Using cases

In the use of cases, teacher candidates either receive or construct context-specific narratives about students, teaching events, or teaching and learning environments; then they analyze and interpret those narratives in the light of other knowledge from research, theory, and experience. When teachers read well-wrought cases, they may better understand certain principles or prototypic dilemmas of teaching that are consciously embedded in the case and made analyzable by the variables and events that are presented. When teachers or teacher candidates construct cases themselves, the writing of the case helps the writer learn to move between levels of abstraction: to understand the relationship between concrete details and larger principles or issues. The opportunity for this to happen is enhanced if there is an interactive process of review and commentary that pushes the writer to explore the deeper meanings of the case and its relationship to other knowledge in the field.

As Shulman (1992, 1996) describes, the initial experience of the case writer is a first-order experience with all the power of any intense encounter with reality. A first-order experience, however, lacks distanced examination and does not necessarily relate experience in a way that easily yields its meaning. As the case writer encodes the experience into narrative, the process requires reflection about what occurred. A process of review and revision, particularly when the reviewer(s) brings broader expertise to bear and when the revision process calls upon research and theory that can illuminate the problematics of the case, explores and elaborates the meanings of the event. When written and shared, the narrative product becomes a second-order experience in two ways. First, it is no longer the experience itself but rather a reconstruction of the experience in language; second, once it is recorded in language, the "experience" become available to a community of peers and colleagues. At the point it is disseminated, the case becomes a third-order experience, because the meaning of the narrative now resides in the community. In this way, as they do in law, business, medicine, and other professions, cases bridge the gap between personal situated knowledge and sharable, generalizable knowledge. The assessment of a case created by a teacher rests on the ability of the case-writer to connect information about teaching events, students, or situations to a broader body of knowledge about learning, teaching, development, culture, motivation, behavior and contextual influences on all of these.

This ideal process may not always occur, thus limiting the usefulness of cases for either teaching or assessing a candidate's abilities to make sense of teaching events or to share them in a manner that has broader professional utility. There are at least two potential dangers. First, a case writer's or case user's limited knowledge or narrow frame of reference may lead to an analysis that misdiagnoses the nature of the problem, misunderstands or ignores the salient context variables, or fails to recognize potentially productive strategies for "solving" or addressing the issues raised in the case. Second, the case writer or case user may be unable to generalize from the single instance represented in the case to a well-grounded set of principles for interpretation or practice, either because he or she is unfamiliar with a broader knowledge base or does not understand how to access and apply this knowledge to the situation under study. When the case is used as an instructional tool, a failure to offset these problems through relevant readings and interpretive

discussion, it could be argued, may misinstruct the case user or developer about the meaning of the situation under study or about principles for practice and their application in specific contexts. When a case analysis is used for assessment purposes, the ability of the assessor to discern and apply the relevant interpretive knowledge is critical to a valid assessment process. Without this lens, cases may add up to little more than interesting but uninstructive teaching stories that reinforce uninformed or idiosyncratic practice. The examples below illustrate efforts to construct educative cases that avoid these problems.

3.3. A case analysis of curriculum

One example of this process is the work done by teacher candidates at Stanford University, who develop curriculum cases as part of a course on Principles of Learning for Teaching. While reading and discussing learning theories in relation to teaching strategies, students write a case about a teaching event in which they have encountered difficulty achieving one of their curricular goals with their students. The case is reviewed on two separate occasions by both a peer and an instructor, who pose questions and possible hypotheses, raise issues and suggest other literature that may shed light on the teaching situation under study. The revisions of the case help the student teacher think more deeply about his experience from multiple perspectives and evaluate a number of potential sources of as well as solutions for the teaching dilemma in question.

Shulman (1996, p. 204) describes one of these cases: a narrative drama in three acts written by teacher education student Mark Ellis while he was working in a geometry class trying to help his students understand the concept of "pi". Ellis knew that all his students had already encountered pi in previous classrooms. He soon discovered, however, that his students' understanding of pi was as a memorized set of digits, "an arbitrary constant with no discernible reason other than that some Greek said several thousand years ago". Mark wanted his students to understand that pi is instead a ratio that is based on the universal, unchanging relationship among the circumference, diameter,

and area of a circle. To help them understand ratio and proportion, Mark designed demonstrations and discussions of scale models, architectural drawings, maps, and other artifacts in which the ideas of scale and proportionality are central. The first act thus consists of an analysis of the complexity of the subject matter concepts as well as of instructional reasoning and strategies.

In the second act, Mark's narrative describes how his plans played out in the classroom. Sometimes the students seemed not to understand even very rudimentary ideas. At other times, his examples and exercises seemed magical. He wondered if his students really understood or were exceptionally adept at framing comprehension. When students completed their final essay examination that included an open-ended question on the meaning of pi, he anxiously awaited the results. Studying their answers, his heart sank. Only two responses reflected more than a superficial understanding of pi.

The third act does not tie up the loose ends like a thirty-minute situation comedy. Mark did not re-teach pi and his students live happily ever after. Rather he carefully analyzed student responses, reflected upon his assumptions and anticipations, and developed a theory that accounted for his experience. His theory was that the persistence of the students' prior knowledge of pi was greater than he anticipated. From that theory he suggested alternative strategies he predicted would be more likely to modify prior misconceptions. As a tool for his development, the case allowed Mark to examine his experience in light of theories of learning and performance. Other novice and veteran teachers may now also use it as an example of the effects of preconceptions on learning. As a tool for assessment, the case allowed Mark's teachers to examine his ability to bring together his understanding of the influence of prior knowledge on student understanding and his personal reflections on his own and his students' intentions and actions.

3.4. A case study of a child

Another approach is the development of case studies in which the author is not the main protagonist or actor but functions instead as a

researcher. The case study can focus on a classroom or school situation or on a particular student. When students are the objects of inquiry, the case study can help teachers learn to apply knowledge of development, learning, motivation, and behavior to specific children as they function in their family, school, and community contexts. Teacher education programs like Bank Street College (see Darling-Hammond & Macdonald, 1999) and Columbia University's Teachers College engage their students in conducting child and adolescent case studies to help them link theories of learning and development to observation of actual children. The goal of such case studies is to examine student learning and development with an eye toward identifying strengths, developmental progress, important influences, and needs. Collecting and analyzing data for the case study – from observations, interviews, records, and analyses of student work – helps teachers develop their skills of observation and documentation and their ability to analyze how children learn and how specific children can be supported in the process of development.

In this kind of case, the narrative explicates with detailed examples a young person's thinking, learning, interactions, beliefs, concerns, and aspirations. The plot is biographical, creating a theory of the person rather than of an event. In some instances, child case studies can be the basis for evaluating how better to work with a child who is having difficulty. Written versions of such studies codify what is done by teachers in action when they individually evaluate a student using multiple tools of evidence, or when teachers collectively participate in a descriptive review of a child in which they pool their observations to figure out how better to support him or her. Like case conferences in medicine, these reviews are build on careful, detailed observation and shared expertise aimed at more powerful analysis of a situation.

A vivid example of this kind of analytic child case study is provided in the account of Akeem, a third grade student who entered Susan Gordon's classroom in a New York City elementary school after having been expelled for throwing a desk at a teacher in another school (Darling-Hammond, Ancess & Falk, 1995, pp. 217–224). The case begins by describing Akeem's frequent outbursts, his ef-

forts to disrupt classroom meetings, and his periodically surly or aggressive behavior. It continues by describing Gordon's efforts to document, using many tools of observation and assessment, exactly when these outbursts occurred, and her discovery that Akeem's misbehavior tended to occur when certain kinds of academic tasks arose. She concluded that Akeem's actions seemed designed to deflect attention from the fact that he could not read well or write with any ease. The case provides a detailed description of Susan's efforts, with her colleagues, to discover what Akeem does do well, to provide opportunities for him to build upon his strengths, and to develop strategies for addressing his specific literacy needs.

Susan allowed Akeem to work in hands-on learning centers that tapped his artistic skills and his abilities to construct machines and models. She found him books and developed writing assignments that built on these interests, while systematically teaching him new strategies for reading. As the case unfolds, Akeem develops architectural drawings and sophisticated comic books which he later annotates and turns into books; he is recognized by peers for his artistic and mechanical abilities and begins to gain status in the classroom; he joins classroom activities with increasing enthusiasm; and, not incidentally, he learns to read and write. The case follows Akeem until he finishes middle school with a solid academic record, near perfect attendance, and admission into a specialized high school for the arts.

This kind of case provides novices with an illustration of how to collect evidence about students' learning and behavior in light of broader knowledge about both; how to diagnose learning needs; and how to build a set of teaching strategies that addresses these needs. When novices construct their own case studies of children, they engage in similar kinds of diagnostic thinking and in an integration of information from many perspectives: cognitive, social, emotional, and physical. Such case analyses are most powerful when they are directly tied to the formal study of development, so that interpretation is guided by the literature and students have a basis for understanding what they are seeing. Even if the story line is not as dramatic, the case construction process enables novices to learn

how to apply theoretical knowledge to concrete examples, and the case provides a basis for evaluating their ability to do so.

3.5. A case analysis of teaching decisions in context

A third approach uses cases to highlight issues of teaching context. For example, the Teachers for Alaska Program at the University of Alaska, Fairbanks uses cases to explore concerns of multicultural teaching in local contexts. Cases in this program offer students a preview of situations they may encounter in their teaching careers; provide descriptions of strategies successful teachers use in handling these special situations; and help novices develop tools for handling the "messy dilemmas that require all the imagination, intellectual resources, and tact at a teacher's command" (Kleinfeld, 1998, p. 145).

Teachers for Alaska replaced the traditional sequence of foundation courses and methods courses followed by student teaching with a program organized around curriculum blocks, each of which emphasizes the study of a case that is thematically related to the subject matter being taught. The cases consist of actual situations confronted by teachers in the culturally diverse classrooms and communities of Alaskan villages. They are modeled on the "dilemmas" approach to case method teaching used by the Harvard Business School to prepare practitioners for action in complex and uncertain contexts (Christensen & Hansen, 1987). These cases introduce students to the "tangled issues of teaching in remote villages – the simmering animosities between local people and high-paid outsiders, unfamiliar cultural rules that new teachers could unwittingly violate, the organization of power in village communities, the injustices the educational system has visited on villagers, and also the injustices visited on outside teachers" (Kleinfeld, 1998, p. 142).

The teaching cases consist of two parts. Part one poses the dramatic problem nested within a web of related issues. For instance, one case begins with a classroom fight between an Eskimo student and an Anglo student. As the case develops, the teacher realizes the fight is related to the Anglo student's cutting remark about the Eskimo students' work

("D minus, huh?"). Later the teacher discovers that the Eskimo student's interpretation of this remark cuts to the core of his identity. "He thinks I am dumb because I am native". The case also develops such contextual issues as the stress of culture fatigue, the hostility of the local community, and lack of support from school administration. In the case, the teacher considers such pedagogical and ethical issues as: What is a fair grading system in an English class where some students are children of "outside professionals" and native speakers of English while others are Yup'ik-speaking children of subsistence hunters? What alternative grading options might be considered? How can competence be supported and recognized? The case presents teaching problems not as prepackaged, neatly solved exercises but as difficult issues to be explored. The critical task for students in their discussion of the cases, as in teaching, is first to understand the range of considerations and from that understanding to determine what to tackle and what to ignore.

Part two of the case shows how experienced teachers go about addressing the issues raised in the first part. The advantage of part two is that it features specific strategies that candidates can consider using themselves. In the above example, for instance, the teacher revises his grading system with individualized goals for each student and grades students on their success in meeting their individualized goals. He creates a bulletin board entitled "The Theme is Excellence" to post students' work, pictures of them doing homework, and articles about their parents from the local newspaper. With other teacher colleagues, he organizes a community relations campaign with a poster showing an Eskimo mother with a baby superimposed over a classroom of students. The caption reads, "WE TEACH … the children you love". In keeping with the program's keep-it-messy theme, this case ends with a paradox: the teacher "burns out" and leaves the community. During the program, students write a case from their own student teaching experience. Many of these cases become part of the curriculum for the program the following year. They provide a base for assessing students' success at understanding their work in a multicultural context and for developing productive strategies for reaching their students.

The case-based approach, coupled with carefully structured coursework and clinical experiences, appears to make a difference for candidates' learning. Evaluations of the Teachers for Alaska Program show measurable improvement in students' cross-cultural teaching skills from the point of entry until graduation. In a recent evaluation, at the end of their first semester on campus and again after their student teaching experiences in the villages, trained observers documented sample lessons graduates taught with culturally diverse students. At their entry into the program, 28% of the candidates took into account culturally different students' frame of reference. At the program midpoint, 62% did so. At the end of the program, 83% did so. At program inception, 12% of the candidates took into account the students' vocabulary and speech patterns, while at the program midpoint, 31% did so, and at the program end, 46% did so. Other measures – such as the use of active teaching strategies rather than lecturing – showed similar changes (Kleinfeld, 1998). Examining and practicing teaching in cultural and community context may indeed strengthen teachers' abilities to take account of their students.

While such cases may be extremely valuable, they need to be constructed with great care, so that they avoid the risk of inadvertently stereotyping students or situations, or attributing to cultural or other characteristics of students behaviors that may have other origins. An example of a case that may illustrate this problem is one written to describe an encounter between a young, white, untrained first year teacher with a group of black male students in her class who did not obey her directives (Skolnick, 1995). Characterized by the case writer as a problem of gender and authority, the encounter could as easily have emerged because of the teacher's pedagogical fumbling, her use of an authoritarian discipline plan that relied on confrontation rather than mutually negotiated goals, her failure to see or address the relationship between one of the student's poor reading skills and his behavior when called upon to read, or her own race-based insecurities with the students in her class. Thus a case that seems to suggest firmer disciplinary action as the answer to a troubling classroom situation may deflect student teachers' attention from important pedagogical concerns. One strategy for vetting cases to test the range of interpretations is to seek out experts with diverse specializations and perspectives to evaluate the cases and write commentaries about them. This strategy has been used in recent casebooks (see e.g. Shulman & Lotan, 1998) to increase confidence in the case interpretations.

4. Exhibitions of performance

While cases provide sites for the analysis of teaching decisions and outcomes, exhibitions of performance directly address the problem of enactment. Exhibitions allow teachers to demonstrate particular abilities in ways that include or closely simulate teaching contexts or events. Exhibitions can draw upon tools such as observations or videotapes of teaching, artifacts like teaching plans, or even group activities that simulate what teachers do when solving problems of practice with colleagues. The distinguishing feature of an exhibition that differentiates it from an unguided observation of practice is that it allows the evaluation of these abilities in relation to articulated standards of practice.

4.1. Defining standards of practice

The teacher education program at Alverno College[2] uses frequent exhibitions of performance, benchmarked against standards, as the foundation for much of its work. The goals of the program are expressed by overarching themes that treat the developmental needs of learners; an appreciation for diversity; a view of professionalism that includes ongoing inquiry to inform teaching; a concern for democratic education; and a commitment to the use of media and technology. The College's curriculum is built upon opportunities for students to master a set of eight general education abilities (expected of all students in the college) and five advanced education abilities (specific to teacher

---

[2] The following discussion draws primarily upon a study of Alverno's teacher education program by Ken Zeichner (1999) and upon Alverno College program documents.

education students). Alverno's definition of abilities includes "a complex integration of knowledge, behaviors, skills, values, attitudes, and self-perceptions" (Diez, Rickard & Lake, 1994, p. 9). The general education abilities include: *communication; analysis; problem solving; valuing in decision making* (the ability to understand the moral dimensions of decisions and to accept responsibility for the consequences of actions taken); *social interaction* (the ability to get things done in groups); *global perspectives* (the ability to understand diverse opinions, ideas, and beliefs about global issues); *effective citizenship* (the ability to engage collaboratively in community issues); and *aesthetic responsiveness* (the ability to make meaning out of artistic experiences and to explain choices of aesthetic expressions). The five professional education abilities include: *conceptualization* (the ability to integrate content knowledge and understanding of education to plan and implement instruction); *diagnosis; coordination* (the ability to manage resources effectively to accomplish learning goals); verbal, nonverbal and media *communication; and integrative interaction* (the ability to act as a professional decisionmaker).

These abilities, as they are expressed according to six levels of increasingly sophisticated attainment, clearly state what program graduates are expected to know, be like, and be able to do to complete the program successfully and be certified as teachers. In general, these levels begin with the ability to identify particular skills or behaviors, and they progress through the abilities to analyze, evaluate, and demonstrate those skills and behaviors, and ultimately, to help others acquire and use these skills in group settings and interpersonal relationships. For example, a beginning teacher in the area of conceptualization is expected to develop sensitivity to individual learners; make links between developmental theory and concrete individuals; plan instruction that recognizes the impact of differences (in culture, gender, learning preferences, etc.) on individuals and the group; and plan material to meet learners' current needs and to lead to the next level of development (Alverno College, 1995).

Faculty believe that the best way to determine how well candidates have developed the abilities is to assess behaviors that are associated with them.

They also believe that such assessments enhance learning by providing feedback on learner strengths and weaknesses and by supporting self-assessment. They developed an elaborate performance-based assessment system that enables candidates and their teacher educators to know how well the candidates can apply their knowledge and skills in realistic contexts. From their very first day at Alverno when they make a videotape of themselves giving a short speech to their peers (a task which will be repeated and re-evaluated later), students are constantly assessed in relation to these abilities. Alverno premises its tight coupling of content, pedagogy, and assessment on the belief that learning occurs best when learners have a good sense of why they are learning something, awareness of the specific standards that they must meet to accomplish this learning, and a way of seeing what they have learned.

4.2. Integrating exhibitions with curriculum

When the college moved to the ability-based curriculum, faculty redesigned all coursework, field experiences, and assessments to insure the systematic development of the knowledge, skills, dispositions, and attitudes implied by the abilities. All of the course syllabi at Alverno spell out which developmental levels of which abilities they address. In addition, syllabi describe the learning activities and assessments that are provided to help students learn the abilities and to judge how well they have learned them. These include essays, letters, position papers, case study analyses, observations of events, talks to simulated audiences, productions of videos and curriculum materials, simulated events such as parent–teacher conferences, and the like.

Students also experience a series of required assessments enabling them to pass from one stage of the program to another. For example, in a fifth semester external assessment integrating learning from several courses, five to six students are asked to take on the role of a teacher group called by the district to review the district's mission statement. Candidates study background materials such as the district's philosophy as well as readings on such issues as curriculum, integration and multicultural education. In preparation for the videotaped as-

sessment, students review the criteria for the abilities of social interaction and effective citizenship by which their performance will be assessed. Following the simulation and before receiving feedback from faculty, they view the tape of the meeting and complete a self-assessment response form (Zeichner, 1999). Virtually every assignment and assessment begins with reference to the criteria for the performance being developed and ends with an opportunity for candidates to evaluate their own work. The end result, as indicated by the judgements of cooperating teachers, college supervisors, employing principals, and candidate assessments of their preparation, is a set of graduates who are both extraordinarily self-reflective and practically well-prepared for sophisticated practice in the classroom (Darling-Hammond, 2000; Zeichner, 1999). As one principal who hires Alverno graduates observed:

> They constantly reflect on their instruction and they're very open to suggestions or to changing a lesson. They're very able to assess the actual lesson they've taught and in a fairly critical manner. They have the skills to do that …. That's not to say that other students are not able to pick it up. It's just that Alverno students seem to come with that knowledge. They've been forced to practice it on an ongoing basis so they have refined it (Zeichner, 1999).

In teacher education, as in elementary and secondary education, there are several potential benefits of continual public practice and assessment through exhibitions of performance. As Ted Sizer (1992) observes of exhibitions in his work with reforming secondary schools, these demonstrations can help make clear what students should be able to do and focus effort accordingly, help faculty "map backward" from their conception of desired learning to a curriculum that can develop such learning, and provide a basis for accountability to the student and to the broader publics that a program serves.

As with other strategies, these benefits are not automatic. They depend on choosing tasks that represent important skills and abilities and on integrating such assessments into a well-developed set of learning experiences. Exhibitions also require

a clear and focused set of goals that adequately reflect the complexities of the tasks to be accomplished and consider their outcomes for teaching and learning. To accomplish this, they demand assessors who are themselves expert in the areas of work being developed. Absent these characteristics, exhibitions can be merely performances without standards, unfocussed activities that provide little guidance or evaluation for developing high levels of skill. The fate of microteaching and competency-based teacher education as used by many teacher education programs in the 1970s comes to mind as one illustration of this potential problem. Although well-developed in some institutions (Alverno's ability-based curriculum was, for example, an outgrowth of the competency movement of that time), in others the techniques of specifying and observing performances became disconnected from a well-grounded base of theory and from standards for performance. Consequently, the exhibitions of behavior often became little more than that – performances in which behaviors were observed and sometimes tallied but not well-assessed in terms of their appropriateness for a particular purpose or context or for their actual or likely influences on students' learning. Assessments that overcome these difficulties integrate exhibitions with a well-conceptualized set of standards, analysis of teachers' goals, contexts, and intentions, and a view of the big picture of teaching and learning for both prospective teachers and their students.

5. Portfolios

The benefits of exhibitions can be further expanded when evidence of performance is assembled to allow a more comprehensive and holistic examination of abilities. Portfolios are means by which teachers select and reflect upon artifacts of their practice collected over time and from multiple sources and diverse contexts to provide evidence of their thinking, learning, and performance. Portfolios can include documents that derive directly from teaching – copies of lesson or unit plans, syllabi, handouts given to students, assignments, tests, and samples of student work (with or without

teacher feedback) – as well as photographs, video-tapes, or audiotapes or classroom activities ranging from bulletin boards and displays, to taped lessons, conferences with students, and the like (Darling-Hammond et al., 1998). They can also include documents that require additional analysis on the part of the teacher, such as teacher logs or journals, detailed descriptions or analyses of lessons or student work, and reflection on the outcomes of teaching activities. Portfolios can include documents that derive from the evaluations of others: notes by an observer of teaching, peer or administrator recommendations, student evaluations, and so on (Bird & King, 1990; Athanases, 1994; Haertel, 1991).

Teacher portfolios provide opportunities for robust documentation of practice. As an assessment tool, they can provide a comprehensive look at how the various aspects of a teacher's practice – planning, instruction, assessment, curriculum design, and communications with peers and parents – come together. As a tool for learning and reflection, portfolios can alleviate what Lee Shulman has referred to as "pedagogical amnesia" – a disease endemic to teaching at all levels. Pedagogical amnesia – characterized by the inability to record and recall the fruits of teaching experience – is actually a symptom of the multidimensional complexity of teaching. So much happens so fast that it is a blur. Portfolios, like cases, help make teaching stand still long enough to be examined, shared, and learned from.

5.1. Approaches to portfolio assessment

In the portfolio assessments of the National Board for Professional Teaching Standards, candidates include collections of the work of several different students over many weeks of teaching, rather like mini-student portfolios. They show and discuss their teaching, evidence of student learning, the feedback they have given to student work, and students' responses to these teaching efforts. In short, they exhibit and reflect upon a set of reciprocal teaching and learning interchanges in which their own learning about their students is as fundamental to the act of teaching as is their students' learning in response to specific lessons. Teacher

and student learning influence each other and are completely interwoven. Teachers consistently testify that the process of developing such a portfolio is a powerful occasion for their own learning (Athanases, 1994; Bradley, 1994; Haynes, 1995; Tracz, Sienty & Mata, 1994; Tracz et al., 1995).

Preservice teacher preparation programs are increasingly using portfolios as means for aggregating and integrating learning experiences, and assessing students' readiness to assume the responsibility of teaching. At Alverno College, performance assessments drawn from exhibitions and other sources are assembled in a portfolio that provides the basis for a portfolio interview assessment that occurs at the end of the preprofessional stage of the program and is used as a gateway to student teaching. Students create a portfolio by reviewing their work in all of their courses to date. They collect examples of written work, lesson and unit plans, videotapes of their work with pupils, and instructional materials they have created and make decisions about what represents their strengths. The portfolio includes a written analysis of a videotaped lesson based on the five abilities. Student faculty advisors as well as teams of school-based educators from area schools review the portfolios. The principal and teacher assessors provide feedback to students about their areas of strength and areas of needed growth as demonstrated in the portfolio and make a recommendation to Alverno about the students' readiness for student teaching. Candidates, using the input of school and college-based assessors, formulate specific goals for their student teaching experience (Zeichner, 1999).

At the University of Southern Maine (USM), secondary teaching candidates develop a portfolio of their practice over the course of a year-long graduate level program in which they are placed as interns in school classrooms while simultaneously completing coursework in teaching methods, learning and development, curriculum, and assessment. The portfolio includes the evidence that candidates offer as a basis for the judgement as to whether they are ready to complete the program and become certified to teach. A panel of university- and school-based faculty makes the ultimate judgement about certification following a portfolio interview in which the candidate presents and defends his or

her work. The portfolio construction process is designed to foster continuous self-reaction and internalization of a set of standards for teaching (Whitford et al., 1999). The standards for what beginning teachers should know and be able to do were developed by faculty drawing upon the model licensing standards offered by the Interstate New Teacher Assessment and Support Consortium (INTASC, 1992), and include statements of knowledge, skills, and dispositions in the following areas:

- knowledge of child and adolescent development and principles of learning;
- knowledge of subject matter and how to make it accessible to students while fostering independent inquiry;
- instructional planning based on knowledge of the learner, the subject matter, the community, the intended student outcomes, and the curriculum;
- uses of instructional strategies and technology to promote learning and independent inquiry;
- assessment for communicating feedback and promoting self-evaluation;
- respect for diversity and the ability to create instructional opportunities for diverse learners;
- well-articulated beliefs about teaching, learning, and education linked to demonstrable practices in support of those beliefs;
- the ability to plan instruction that promotes the values and practices of citizenship;
- the ability to work collaboratively to improve the conditions of learning for students and adults;
- commitment to reflection and continuous professional development; and
- classroom management that supports individual responsibility and democratic community.

Candidates assemble a body of evidence to demonstrate their learning and competence across these areas and may include artifacts like their own statements of teaching philosophy, classroom lessons, student work, and so on. The process of portfolio construction also includes critical conversations regarding the candidate's practice with mentors or peers who are part of a portfolio team and reflections on the contents of the portfolio that describe what each entry represents and why it is included, what the teacher learned from the experience about teaching and learning, and why that is important – that is, the personal meaning of the learning. The process concludes with a portfolio presentation to faculty and peers that is a major aspect of the final graduation and certification decision (Lyons, 1998, p. 19).

5.2. Outcomes of a portfolio process

This process of construction and reflection upon the portfolio is as important for candidate learning as are the components of the portfolio itself. It is through this process of selecting and discussing artifacts of their practice that candidates internalize the standards, examine more deeply what they are doing and what it means, and gain multiple perspectives on the meaning of events, thus enhancing their ability to learn from those events (Whitford et al., 1999). This notion is implicit in Lee Shulman's (1994) definition of a teaching portfolio as "the structured documentary history of a (carefully selected) set of coached or mentored accomplishments substantiated by samples of student work and fully realized only through reflective writing, deliberation, and serious conversation". The ways in which these processes contribute to the value of candidates' learning are illustrated by these comments from two of USM's teacher interns:

> The portfolio process worked best when it helped me to reflect on what I can improve on and what I did well. I have found that when I am in the process of something I become convinced that I will remember it – including all the details. But I don't .... I have found that when I reflect and create my portfolio, I reinforce the event in my memory so that I am less likely to forget about it. The process of reflection imprints the event in a unique way .... What this points to is the need to create portfolios contemporaneously with the [teaching] process. (Davis & Honan, 1998, p. 96).

> Designing my portfolio helped me to clarify and articulate visually and in writing my teaching philosophy. It is the actual process that I value: collecting artifacts, organizing, reflecting, and receiving feedback at the portfolio evaluation ....

By receiving positive feedback and constructive criticism, I concluded that my portfolio is a continuum of my learning as a teacher, i.e., it will *never* be done. Instead of my portfolio presentation being a final, pass/fail assessment, it was a learning experience in and of itself. (Davis & Honan, 1998, p. 98)

The benefits for teacher learning of well-constructed portfolios seem related to their ability to:

- Raise teaching decisions to consciousness and thus make them available for deeper consideration from many perspectives. The process of looking at and thinking about decisions changes consciousness about teaching, and thus changes practice. Beginning teacher candidates who undergo such forms of assessment have to answer the question, "What am I aiming for as I learn to teach?"
- Take a long view of learning and of the development of performance. Because proficient performance must be developed over a long period of time with continuous practice and reflection on practice, a cumulative record helps both to scaffold and evaluate that process.
- Support the developmental process by providing benchmarks for good work, vehicles for self-assessment and peer assessment, and opportunities for revision and refinement.
- Connect thinking and performance. This helps to develop the capacity for reflection and action, rather than just one or the other. They bridge the traditional theory–practice divide by asking for evidence of performance along with a discussion of why decisions and actions were taken.
- Provide multiple lenses and multiple sources of evidence on thinking and performance, thus developing many facets of performance and allowing many pathways into learning
- Make teaching and learning more public, thereby making the development of shared norms and standards possible, as well as making the sharing of knowledge and experience more available.

These factors combine to enhance the candidates' abilities to integrate the knowledge, skills, and dispositions required of teaching and to provide tools for continuous development once teaching. As veteran teacher Shirley Bzdewka observed after completing the NBPTS portfolio process:

I am a very different teacher now. I know I was a good teacher. But I also know that every teacher always has a responsibility to be better tomorrow than they were today, and I am a much more deliberate teacher now. I am much more focused. I can never, ever, do anything again with my kids and not ask myself, 'Why? Why am I doing this? What are the effects on my kids? What are the benefits to my kids?' It is not that I didn't care about those things before, but it is on such a conscious level now. (NCTAF, 1996).

As assessment tools, portfolios that are structured around standards of practice are able to examine a teacher's practice both in context and in the light of a common set of expectations and benchmarks. By giving assessors access to teachers' thinking as well as to evidence of their behaviors and actions (e.g. through videotapes, lesson plans, assignments, and the like), portfolios permit the examination of teacher deliberation, along with the outcomes of that deliberation in teacher's actions and student learning. The long-range view that is encouraged and supported by portfolios helps assessors overcome some of the "limits of looking" (Stodolsky, 1984) that have plagued traditional observations of teaching. Assessors can examine a chain of events and thinking, analyze the quality of deliberation and the grounding of decisions, evaluate the quality and appropriateness of actions taken, and take into account the evidence of student characteristics and learning that are a basis for gauging effectiveness. In short, assessors using portfolios can "see" teaching in progress and make it hold still long enough to understand its intentions and effects.

As with other strategies, these benefits are not inevitable. Portfolios, like cases, are a step away from teaching itself, and can privilege a candidate's ability to select and write about artifacts of teaching more than the candidate's capacity actually to teach well in the classroom in the heat of real situations with real students. It is undoubtedly important that assessors of portfolios both be expert in the field and savvy enough to sort out evidence of solid teaching from the public relations

trappings that could characterize such collections. Furthermore, it portfolios are not organized around a conception of teaching knowledge and skill and assessed in ways that represent that conception well, they can be mere collections of assignments bound together in a folder. Portfolios are likely to be more useful for developing teachers' insights and judgments if they encourage reflection and revision of work against well-grounded standards of practice in the field. If this kind of assessment is to shape novices' thinking about teaching, candidates and assessors must know what they are aiming for, what criteria for "good teaching" will be applied in the review of work, and how each element of the portfolio contributes toward the overall picture of teaching.

6. Problem-based inquiries

Yet another way to promote deeper understanding of teaching in context is to embed systematic research about the contexts and outcomes of teaching and schooling into candidates' programs of study. In action research or inquiry projects, teachers design and conduct investigations into concerns arising from their work with children and families. These inquiries may involve questions similar to those that teachers may explore with cases, but the method extends beyond personal reflection about an individual's experiences and observations to broader and more structured investigation involving the collection and aggregation of data and information about a problem. Teacher research advocate Marilyn Cochran-Smith (1991) suggests that the posing and pursuit of questions provide important vehicles for teachers to understand both the complexities of teaching and the effects of different solutions or resolutions of endemic problems. "The ability to pose questions", she argues, "to struggle with uncertainty and build evidence for reasoning … is an indispensable resource in the education of teachers" (pp. 280–281).

Teacher research – like cases, exhibitions, and portfolios – can transform teaching from a private and hidden act into community property (Shulman, 1996). The sharing and critique of practice, along with research related to practice, create the corner-

stones of professions. When teaching is treated as community property, problems, conjectures, analyses, and interpretations can be examined by collaborating professionals. These inquiries can be preserved for future study and can be drawn upon and built upon by others. Such knowledge construction is not solely the domain of the outside "expert". Beginning teachers come with their own perspectives and interests. Having prospective teachers engage from the start of their careers in action research or classroom inquiry can help prepare them both as consumers of research and producers of knowledge. It can also give them tools to make sense of their practice and help them think analytically about the problems they confront. As a tool for assessment, such research and inquiry can provide insights into a teacher's analytic ability and her ability to frame a problem in a manner that allows it to be thoughtfully examined (Lippincott, 1999). Such studies can reveal a teacher's disposition and skills for responding to problems of practice with action-oriented hypotheses that may lead to improvement, rather than merely copying.

At the University of California at Santa Barbara, student teachers develop an action research project as the culminating assessment for the Masters program of study. Called the "M.Ed. portfolio", the inquiry is reported through a collection of studies and reflections about an issue of practice developed over the course of at least eleven months.[3] Early in their professional preparation year (August), students take an ethnography course to help them learn to collect data in natural settings. By December, most students have identified an inchoate passion about some element of teaching and learning. Through a series of workshops and field-based experiences, students refine and focus their questions, moving between questioning and reflection upon the concrete artifacts they have been collecting in their field experiences and in their coursework. The data they collect may include articles from the research literature and other readings on the topic, analyses they have conducted through research papers and through data collected in their

---

[3] This discussion draws upon Snyder, Lippincott & Bower (1998).

school or classroom, reflections on observed events that bear on the topic, and examples of their own teaching efforts and outcomes that bear upon the topic.

By March students form self-selected support groups and are assigned a facilitator. These support groups meet regularly so that members can inform one another of their thinking/practice regarding *their* issue along with the evidence they have selected to document the outcomes of their inquiry and their learning and growth over time. These conversations, like those that accompany the construction of other kinds of portfolios, provide multiple perspectives on the topic, raise new questions, and provoke deeper thinking. In the summer following their year-long full-time student teaching experience, students complete the M.Ed. Portfolio.[4] Recent examples of inquiry topics include:

- What is inclusion? In what ways has my teaching/the system created inclusion/exclusion?
- How can I make my US History curriculum meaningful to my second-language learners?
- What are the roles of modeling in creating clear, explicit teaching while simultaneously allowing for creative and original student work?
- In what ways do I/can I assess transition ESL students?
- What do students really learn from doing homework?

Successful completion of the project involves two checkpoints. First, the group facilitator and every member of the support group must give his or her approval to the document. Once approved by the group, students schedule a public conversation where they receive feedback on their work from five critical friends. These include a school-based educator who knows the candidate well (i.e., a cooperating teacher), a school-based educator who does not know the candidate well (i.e., a principal, another teacher), a university-based educator who knows the candidate well (i.e., the supervisor), a university-based educator who does not know the candidate well (i.e., a content expert or researcher), and someone whose primary intersection with the school is as a parent or in a community/social service capacity.

Several of these public conversations occur simultaneously in a large room, somewhat like a poster session at the American Educational Research Association. Critical friends arrive prior to the session and review the entire document without the candidate present. The conference then is not a two-hour presentation of the work, but rather a two-hour conversation about teaching and learning among professional educators about a topic of mutual concern. Projects are assessed on their coherence and clarity of ideas; significance of the topic for the field; evidence of growth over time in the way the candidate frames and understands the research questions, changes techniques and/or attitudes, and learns from problems and successes; and implications for future development in teaching and learning.

Like a dissertation, one goal of the inquiry and its assessment is to develop and evaluate skills of investigation and analysis. In addition, the project is structured to encourage direct applications to practice. Finally, the process of evaluation is organized to ensure multiple perspectives on the question, including those of parents or community members, and feedback from various sources. The goal is the development of a thoughtful practitioner who has tools to inquire into and address problems of practice throughout his or her career. This goal may or may not be achieved by the inquiry strategy, depending on how well the guidance offered enables students to learn to generate useful questions and approaches to study and how to examine and make sense of the results. If this guidance is adequate, the habit of looking at the outcomes of teaching strategies as a researcher could shape a long-term view of practice.

At UC-Santa Barbara, the emphasis on assessment of one's own learning as well as that of

---

[4] Some students postpone completion of the M.Ed. portfolio, electing to give themselves another year in the belief that additional experience and reflection will enrich their portfolio, their teaching, and the learning of their students. The program holds a series of Saturday workshops through the subsequent year to support the growth of these students. Despite the logistical problems and the unpaid time and labor demands of this model, it remains the preferred choice of the program faculty who work with these first-year teachers.

students appears to shape teachers' later work. In a school of education follow-up study of graduates, first-year teachers reported that they were using assessment tools with children that their teacher educators used (Snyder et al., 1998). The portfolio appeared to help teachers connect their self-assessment with their assessment practices with students. In addition, UCSB program graduates rated themselves better prepared to assess student learning and to use their knowledge of student learning to shape instruction than did a random national sample of beginning teachers (Snyder, 1997).

7. Side-effects of authentic assessment practices

The use of authentic assessment practices in teacher education appears to hold potential, at least under some circumstances, for influencing the learning of teachers and thereby the learning opportunities of their students. In addition, the use of such practices appears to support ongoing program improvement. Each of the sites described above has conducted studies of the effects of their work which found evidence that the information provided by these assessments provoked programmatic changes. A study of the University of California – Santa Barbara program found, for instance, that "portfolios made our program more visible to us as well as illuminated the developmental nature of teaching" (Snyder et al., 1998, p. 138). Faculty found that the portfolio processes undertaken in this program had the unintended consequence of helping cooperating teachers and others involved in the assessment of student teachers become more thoughtful about their own practice as well as their mentoring of novices. School-based educators gained a greater understanding of the program and of the state's standards for teaching, which created more coherence between school- and university-based work; they also began to construct their own portfolios as a means of enriching their own practice.

Authentic assessment can inform ongoing instruction as well, by revealing what students understand well enough to apply. For instance, when one of the authors taught a course on Adolescent Development, her students' efforts to develop

a case study of an adolescent illuminated the fact that few understood how to examine a student's thinking or how to evaluate cognitive development. They had really never learned how to look for evidence of reasoning and understanding. As a result, she began to treat the issues of cognitive development more extensively and concretely, using videotaped and written cases of student learning and student work samples to illustrate how cognitive development could be examined and assessed. Having more than superficial evidence about the thinking and reasoning of one's own students can allow a teacher educator to better adapt instruction and to redefine program content.

Although anecdotes suggest the promise of these tools for improving practice, the field lacks systematic research evidence linking pre-service teacher learning opportunities with inservice teacher classroom practices, and teachers' classroom practices with student learning outcomes. As Kleinfeld (1998) notes, "Virtually no formal evaluations of case-based teaching have been done to ask if teachers prepared through the case method actually teach better" (p. 145). The same could be said of any of the assessment practices described above. The difficulty in conducting such research may be partly attributed to the multitude of variables that would need to be considered and to the difficulty of drawing inferences about cause and effect given the many factors that influence teacher and student learning. Still, recent work that has established links between specific kinds of teacher learning with teaching practices and student learning (Cohen & Hill, 1997) suggests that such research may be successfully pursued.

The use of these curriculum and assessment practices in teacher education programs also poses challenges that educators and policy makers need to take into account if such practices are to meet their potential for supporting the learning of teachers and their students. As an add-on, authentic assessment of teaching is too expensive in terms of time and money to be feasible. Unless such assessment is embedded within a program of study and treated as part of the curriculum and teaching process, it is not likely to be sustained. Embedding authentic assessment of teaching in preservice teacher education requires the time and expertise of

school-based educators, college-based educators, and prospective educators working closely together to develop and discuss practice over time. These three sets of resources need to be concentrated in time and space, rather than fragmented as is often the case. Such practices may be more feasible when there are groups of prospective educators working with a team of cooperating teachers at a school site, as in a professional development school model, and clusters of teacher education faculty working together with one another and school-based faculty (Snyder, 1998).

Such concentration of effort only works, however, when supportive structures and processes are in place. One of the most important structures is time to develop and sustain the conversations and relationships necessary for the sustained learning activities embodied in authentic assessments. This in turn requires both rethinking of how existing resources are organized and used and how schools and colleges fund and organize responsibility for the education of teachers. Among the policies that are implicated in these matters are funding streams from state governments that typically discourage collaboration between schools and colleges, university hiring policies that often provide too few faculty for the work of teacher education, tenure and promotion policies that fail to reward work with schools, and fragmented curriculum that emphasizes superficial coverage over the development of proficiency and understanding. Unless policies that encourage collaboration and coherence are in place, it is unlikely that programs can enact new assessments of teaching without a significant influx of funds and greater time demands on staff (Snyder, 1998).

Even with these institutional conditions in place, however, teacher educators face a dilemma comparable to the one faced by K-12 teachers who use authentic assessments: What should they do if there are mismatches between assessments mandated by states and embedded authentic assessment practices? Most states, for example, still use multiple choice tests of subject matter and teaching knowledge as the basis for granting a beginning teacher license. (Connecticut's portfolio assessments for beginning teachers are an exception to the norm.) These tests increasingly determine candidates' op-

portunities to teach and the fate of teacher education programs seeking professional accreditation or state approval. Should teacher education programs continue to try to develop and use authentic, contextualized assessments of teaching? Or will they sacrifice their candidates' ability to receive a credential if they do not turn their programs into courses of test preparation? How well will candidates from "authentic" programs perform on the exam(s) as compared to those who attend "test prep" programs? How effective are differently prepared teachers in the classroom? Until teacher educators subject their work to rigorous inquiry that can begin to answer these questions, policy and practice are likely to remain at odds.

Among the questions that are important for future research are, at least, the following:

- How well do different types of assessments measure the capacity to teach? What evidence can be developed of the predictive and consequential validity of various measures?
- What are the effects on teacher learning of the use of different types of assessment?
- Given that no single measure of teaching is adequate to the task of representing such a complex activity, what mix of assessment methods, instruments, and sources of evidence seems to provide the greatest leverage on teacher development, on the one hand, and valid assessment, on the other?

Continued work on these questions may enable teachers and teacher educators to develop strategies that are both powerful and practical for the development and evaluation of contextualized teaching.

References

Alverno College (1995). *Handbook for education students, Part Two: Conceptual frameworks*. Milwaukee, WI: Alverno College.

Andrews, J. W., Blackmon, C. R., & Mackey, A. (1980). Preservice performance and the National Teacher Examinations. *Phi Delta Kappan*, 6(5), 358–359.

Athanases, S. Z. (1994). Teachers' reports of the effects of preparing portfolios of literacy instruction. *Elementary School Journal*, 94(4), 421–439.

Ayers, J., & Qualls, G. (1979). Concurrent and predictive validity of the national teacher examination. *Journal of Educational Research*, *73*(2), 893.

Bliss, T., & Mazur, J. (1997). *How INTASC standards come alive through case studies*. Paper presented at the Annual Meeting of the American Association of Colleges for Teacher Education, Phoenix, AZ.

Bradley, A. (1994). Pioneers in professionalism. *Education Week*, *13*, 18–21.

Christensen, C., & Hansen, A. (1987). *Teaching and the case method*. Boston: Harvard Business School.

Cochran-Smith, M. (1991). Learning to teach against the grain. *Harvard Educational Review*, *61*(3), 279–310.

Cohen, D., & Hill, H. (1997). *Instructional policy and classroom performance: The mathematics reform in California*. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL, March.

Colberg, J., Trimble, K., & Desberg, P. (1996). *The case for education: Contemporary approaches for using case methods*. Boston: Allyn & Bacon.

Darling-Hammond, L. (1986). Teaching knowledge: How do we test it? *American Educator*, 10 (3), 18–21, 46.

Darling-Hammond, L. (1997). *The right to learn*. San Francisco: Jossey-Bass.

Darling-Hammond, L. (1999). *Reshaping teaching policy, preparation, and practice: Influences of the National Board for Professional Teaching Standards*. Washington, D.C.: American Association of Colleges for Teacher Education.

Darling-Hammond, L. (2000). *Educating teachers*. San Francisco: Jossey Bass, forthcoming.

Darling-Hammond, L., Ancess, J., & Falk, B. (1995). *Authentic assessment in action: Studies of schools and students at work*. NY: Teachers College Press.

Darling-Hammond, L., & Macdonald, M. B. (1999). Where there is learning, there is hope: Bank Street College of Education. In L. Darling-Hammond, *Studies of excellence in teacher education: preparation at the graduate level*. New York, Washington, DC: National Commission on Teaching and America's Future and American Association of Colleges of Teacher Education.

Darling-Hammond, L., & Sclan, E. (1992). Policy and supervision. In C. Glickman, *Supervision in transition*. Alexandria, VA: Association for Supervision and Curriculum Development.

Darling-Hammond, L., Wise, A. E., & Klein, S. (1998). *A license to teach: Building a profession for 21st century schools*. San Francisco: Jossey Bass.

Davis, C. L., & Honan, E. (1998). Reflections on the use of teams to support the portfolio process. In N. Lyons, *With portfolio in hand: Validating the new teacher professionalism* (pp. 90–102). New York: Teachers College Press.

Diez, M., Rickard, W., & Lake, K. (1994). Performance assessment in teacher education. In T. Warren, *Promising practices: Teacher education in liberal arts colleges*. Lanham, MD: University Press of America and Association of Independent Liberal Arts Colleges for Teacher Education.

Doyle, W. (1979). Classroom tasks and students' abilities. In P. L. Peterson, & H. J. Walberg, *Research on teaching*. Berkeley, CA: McCutchan.

Floden, R. E., & Klinzing, H. G. (1990). What can research on teacher thinking contribute to teacher preparation? A second opinion. *Educational Researcher*, *19*(5), 15–20.

French, R. L., Hodzkom, D., & Kuligowski, B. (1990). *Teacher evaluation in SREB states. Stage I: Analysis and comparison of evaluation systems*. Paper presented at the Annual Meeting of the American Educational Research Association, Boston, MA.

Fuller, F. (1969). Concerns of teachers: A developmental conceptualization. *American Educational Research Journal*, *6*(2), 207–226.

Garibaldi, A. (1992). Preparing teachers for culturally diverse classrooms. In M. Dilworth, *Diversity in teacher education: New expectation* (pp. 23–39). San Francisco: Jossey-Bass.

Goodlad, J. (1990). *Teachers for our nation's schools*. San Francisco: Jossey Bass.

Haertel, E. (1991). New Forms of Teacher Assessment. In C. Grant, *Review of Research in Education*, 17, 3–29.

Haertel, E. H. (1990). Performance tests, simulations, and other methods. In J. Millman, & L. Darling-Hammond, *The new handbook of teacher evaluation: Assessing elementary and secondary school teachers* (pp. 278–294). San Francisco, CA: Sage.

Haney, W., Madaus, G., & Kreitzer, A. (1987). Charms talismanic: Testing teachers for the improvement of American education. In E. Z. Rothkopf, *Review of Research in Education*, 14, 169–238.

Haynes, D. (1995). One teacher's experience with National Board assessment. *Educational Leadership*, *52*(8), 58–60.

Hollins, E. (1989). *A conceptual framework for selecting instructional approaches and materials for inner-city black youngsters*. Paper commissioned by the California Curriculum Commission, Sacramento, California.

Hoover, N. L., O'Shea, L. J. (1987). *The influence of a criterion checklist on supervisors' and interns' conceptions of teaching*. Paper presented at the Annual Meeting of the American Educational Research Association, Washington, D.C.

Ingvarson, L., & Marrett, M. (1997). Building professional community and supporting teachers as learners: The potential of case methods. In L. Logan, & J. Sachs, *Meeting the challenge of primary schooling for the 1990s*. London: Routledge.

Interstate New Teacher Support and Assessment Consortium (INTASC) (1992). *Model standards for beginning teacher licensing and development: A resource for state dialogue*. Washington, D.C.: Council for Chief State School Officers.

Irvine, J. (1990). *Beyond role models: The influence of black teachers on black students' achievement*. Paper presented at Educational Testing Service, Princeton, NJ, May.

Irvine, J. (1992). Making teacher education culturally responsive. In M. Dilworth, *Diversity in teacher education: New expectation* (pp. 79–92). San Francisco: Jossey-Bass.

Katz, L. (1972). Developmental stages of preschool teachers. *Elementary School Journal*, *23*(1), 50–51.

Kennedy, M. (1999). The role of preservice teacher education. In L. Darling-Hammond, & G. Sykes, *Teaching as the learning profession* (pp. 59–85). San Francisco: Jossey Bass.

Kleinfeld, J. (1998). The use of case studies in preparing teachers for cultural diversity. *Theory into Practice, 37*(2), 140–147.

Koppich, J. (1999). *Trinity University: Preparing teachers for tomorrow's schools*. In L. Darling-Hammond, *Studies of excellence in teacher education: Preparation in a five-year program*. New York, Washington, DC: National Commission on Teaching and America's Future, American Association of Colleges for Teacher Education.

Lippincott, A. (1999). *Reflective thinking among and between beginning professionals*. Unpublished dissertation, University of California, Santa Barbara.

Lyons, N. (1998). *With portfolios in hand: Validating the new teacher professionalism*. New York: Teachers College Press.

Merseth, K., & Koppich, J. E. (1999). Teacher education at the University of Virginia: A study of English and mathematics preparation. In L. Darling-Hammond, *Studies of excellence in teacher education: Preparation in a five-year program*. New York, Washington, DC: National Commission on Teaching and America's Future, American Association of Colleges for Teacher Education.

Miller, L., & Silvernail, D. (1999). Learning to become a teacher: The Wheelock way. In L. Darling-Hammond, *Studies of excellence in teacher education: Preparation in the undergraduate years*. New York, Washington, DC: National Commission on Teaching and America's Future, American Association of Colleges for Teacher Education.

Moll, L. (1988). Some key issues in teaching Latino students. *Language Arts, 65*(5), 465–472.

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

Ortiz, A., & Maldonado-Colon, E. (1986). Reducing inappropriate referrals of language minority students in special education. In A. Willig, & H. Greenberg, *Bilingualism and learning disabilities* (pp. 37–50). New York: American Library Publishing Company.

Quirk, T., Witten, B., & Weinberg, S. (1973). Review of studies of the concurrent and predictive validity of the national teacher examinations. *Review of Educational Research, 43*(1), 89–113.

Shavelson, R., & Dempsey-Atwood, N. (1976). Generalizability of measures of teacher behavior. *Review of Educational Research, 46*, 553–612.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57*(1), 1–22.

Shulman, L. (1992). Toward a pedagogy of cases. In J. Shulman, *Case methods in teacher education* (pp. 1–29). New York: Teachers College Press.

Shulman, L. (1994). Portfolios in historical perspective. Presentation at the Portfolios in Teaching and Teacher Education Conference, Cambridge, MA.

Shulman, L. (1996). Just in case: Reflections on learning from experience. In J. Colbert, K. Trimble, & P. Desberg, *The case for education: Contemporary approaches for using case methods* (pp. 197–217). Boston: Allyn & Bacon.

Shulman, J., & Lotan, R. (1998). *Groupwork in heterogeneous classrooms: A casebook*. San Francisco: Jossey Bass.

Sizer, T. (1992). *Horace's school: Redesigning the American high school*. Boston: Houghton Mifflin.

Skolnick, J. (1995). Diane: Gender, culture, and a crisis of classroom control. In J. S. Kleinfeld, & R. S. Yerian, *Gender tales: Tensions in the schools* (pp. 111–116). NY: St. Martin's Press.

Snyder, J. (1997). *UCSB experimental teacher education program: Year one report*. Document prepared for the California Commission on Teacher Credentialing, September, 1997.

Snyder, J. (1998). *Finance and policy structures that support the sustenance of professional development schools*. Washington DC: NCATE.

Snyder, J. (1999). Knowing children, understanding teaching: The developmental teacher education program, University of California, Berkeley. In L. Darling-Hammond, *Studies of excellence in teacher education: Preparation at the graduate level*. New York, Washington, DC: National Commission on Teaching and America's Future and American Association of Colleges of Teacher Education.

Snyder, J., Lippincott, A., & Bower, D. (1998). Portfolios in teacher education: Technical or transformational? In N. Lyons, *With portfolios in hand: Validating the new teacher professionalism*. New York: Teachers College Press.

Stodolsky, S. S. (1984). Teacher evaluation: The limits of looking. *Educational Researcher, 13*(9), 11–18.

Tikunoff, W. (1985). *Applying significant bilingual instructional features in the classroom*. Rosslyn, VA: National Clearinghouse for Bilingual Education.

Tracz, S. M., Sienty, S., & Mata, S. (1994). *The self-reflection of teachers compiling portfolios for National Certification: Work in progress*. Paper presented at the Annual Meeting of the American Association of Colleges for Teacher Education. Chicago, IL.

Tracz, S. M., Sienty, S. Todorov, K., Snyder, J., Takashima, B., Pensabene, R., Olsen, B., Pauls, L., & Sork, J. (1995). *Improvement in teaching skills: Perspectives from National Board for Professional Teaching Standards field test network candidates*. Paper presented at the annual meeting of the American Educational Research Association. San Francisco, CA.

Villegas, A. (1997). Assessing teacher performance in a diverse society. In L. Goodwin, *Assessment for equity and inclusion: Embracing all our children* (pp. 262–278). New York: Routledge.

Whitford, B. L., Ruscoe, G., & Fickel, L. (1999). Knitting it all together: Collaborative teacher education in Southern Maine. In L. Darling-Hammond, *Studies of excellence in teacher education: Preparation at the graduate level*. New York, Washington, DC: National Commission on Teaching and America's Future and American Association of Colleges of Teacher Education.

Zeichner, K. (1999). Ability-based teacher education: Elementary teacher education at Alverno college. In L. Darling-Hammond, *Studies of excellence in teacher education: Preparation in the undergraduate years*. New York, Washington, DC: National Commission on Teaching and America's Future, American Association of Colleges for Teacher Education. Washington, DC: American Association of Colleges of Teacher Education.