Novel Approaches to Performance Assessment

Patricia Brinlee

I’d like to say, first of all, that it’s very fortunate that Ray Pecheone’s presentation came before mine, because there is considerable overlap in the kind of philosophical and conceptual approaches that our two states are using. We’re part of an interstate consortium that includes other states as well, but California and Connecticut—I do that in alphabetical order—have been more or less the leaders in this generation of authentic teacher assessment, trying to move to performance-based approaches, novel approaches, and learning many lessons together. Some of the lessons that Ray alluded to are similar to what we’ve learned. For example, the critical importance of conceptualizing the scoring process simultaneously with the development of the assessment. And that’s been a problem for, I think, almost every project that has worked in this area.

California’s Teacher Assessments

I’d like to tell you a little bit about the assessments that are in place for our beginning teachers in California, then go to a video presentation about our support program for induction purposes, and finally talk a little bit about some of the assessments that we’ve piloted since the year before last. We have an ongoing test of basic skills—reading, writing, and math—the California Basic Educational Skills Test (CBEST), which was developed in California and is used in California and Oregon. It must be taken prior to entering student teaching, and it’s also required for teachers coming in out of state for most programs, as well as by teachers moving from one credential to another. For example, classroom teachers who go back for a service credential in administration must take and pass the CBEST. It has a compensatory scoring system and has worked well for the state in the past few years.

Patricia Brinlee is on the Commission on Teacher Credentialing for the state of California.
Transitions to Performance Assessments

Multiple and Single Subject Credentials

In California, to become a teacher, just figuring out all the hurdles is probably the test of whether you make it or not, because we have some very different types of arrangements. One is due to the Ryan Act passed in 1970, which led to a total separation between preparation in subject matter programs and in education coursework. It’s a very different approach from that which many other states have. You cannot, for example, major in Elementary Education, or Early Childhood Education, or Secondary Education; you must major in an academic field and get a bachelor’s degree. Then you must take what’s called the “fifth year,” but that can be completed anytime later. You are required to take specific education content, including student teaching during this phase of professional preparation. That has cut back somewhat on the pedagogical training that teachers receive.

The NTE Specialty Area Exam

The way the law operates, the knowledge required in the content area is measured by the NTE Specialty Area Exam in the particular area one is majoring in. However, if a candidate passes that exam (and the standards tend to be much higher than in other states, where it’s an exit requirement for a program), you can waive the course requirements in the subject area. So the reliability and the validity of those tests are very critical. In a validation review in 1988, some problems surfaced with the tests that were being used, and we used sixteen of them. We are now working with advisory panels in these areas and modifying the tests by adding a performance component to them. For example, I worked with the English advisory panel. We met ten to fifteen times during the last year and a half developing standards for evaluating teacher education programs in the subject matter and developing specifications for an add-on portion to the NTEs. This sounds very much like the successor phase that you’re probably familiar with. I think there’s an attempt to dovetail these two, since we are represented on the panels for developing these stage two assessments. We’re also developing a new elementary education test which is tentatively called the Multiple Subjects Test, and it’s to be a content-based test. We did pilot the Connecticut Elementary Certification Test (CONNECT), the one that Jim Popham developed, and we’ve learned quite a bit from
that. We’re developing, with panel input, these new assessments for elementary teachers, to make them more content-specific and job-relevant.

**New Teacher Support and Assessments**

In addition, we have a cooperative program between the State Department and the Commission on Teacher Credentialing, a support and assessment program for beginning teachers. This has been a really intense project. We’ve learned a lot, we’ve done a lot. We’ve completed considerable research on the process of developing assessments and piloting them, so first I’d like to show you what that project looks like in its various forms around the state.

We now have 37 local projects operating. While they focus mainly on support, in this video you can see the formative assessment that’s being done between the mentors and the new teachers, and also some of the ways in which we are able to help retain new teachers in the schools through this project. So I’d like to show you that, and then I’ll talk about the new assessments we’re developing to fit into these models. [Videotape: “Passage Into Teaching”]

I hope the videotape was illuminating about the support aspect of the new teacher project. Even though there was not much direct discussion of the assessment component, I think you saw quite a bit of formative assessment going on there. Another thing that’s quite important about these projects is that they are indigenous to the particular area. In some models you have cooperation between a university and the school system; each one has to have some kind of collaborative work going on in order to be funded. I think that has helped coalesce some of the thinking and work and it’s also had great spin-off effects for the mentor teachers. Another advantage that we found, which Ray was pointing out, was that there was a considerable amount of informal staff development going on. We’re also breaking down that norm of isolation that Dr. Sykes was talking about yesterday. So at this point, I’d like to pass out these handouts and then I’ll talk briefly about the assessments we have piloted and what our plans are for the next phase.
Phase One of the Assessment Component

To begin with, we have phase one of the assessment component, during which we selected from existing assessments those that seemed promising and were developed with the notion of innovation in mind, trying to look at new ways of assessing teachers besides using traditional multiple-choice tests. We did pilot at least three of the Connecticut assessments that Ray Pecheone was talking about earlier. The CCI we piloted. We had some wonderful trainers who came from Connecticut and worked with our people in Los Alamitos with the Southwest Regional Laboratory for Educational Research and Development (SWRL). We used the CCI in two contexts: first as a part of the evaluation of the support, and second as an assessment tool for teachers in the projects. We also piloted the instrument developed by IOX for Connecticut in a preliminary form. We conducted form-equivalency and content-equivalency studies. In all of these studies we have been very careful to protect the candidates’ confidentiality, and to let them know that they are piloting the assessments and that we are not looking at their performance, per se, at this point in time, because we simply don’t have the scoring protocols and reliability yet. We haven’t had an opportunity to get the number of cases we need to establish reliability. We really are trying to find strengths and weaknesses in these assessments so that these can be addressed before we were to recommend that they go on line.

We have also tried to promote the concept that all these are prototypes or models of assessment approaches, and that should we adopt any particular mode of assessment, there would be considerable adaptation and refinement needed before it could be used to make critical decisions such as moving from a preliminary to a clear credential. Even when we pilot test them, the stakes are much lower. So we have much more work to do in that area.

We also piloted some of the assessments developed by the Teacher Assessment Project at Stanford. These assessments were designed for exemplary teachers, and so we had concerns about using them with the beginning teachers. But they did help us to learn more about the structured interview format. From Connecticut we also used some of the semistructured interviews. This year we’re also going to pilot social studies components that Ray was showing earlier. We’ve
learned a great deal about subject-specific pedagogy—how it works, how it can be tested, the frustrations involved in trying to test it, and so forth.

**Phase Two of the Assessment Component**

That was phase one. I’d like to move on now to the assessments we’ve developed that we’re piloting now. If you’ll look at the back of your handout, there is a chart that shows five assessments that we contracted for development last year. I can show you a little bit about them but I won’t go into too much detail and will not have time to look at the videos on them.

**The Videotaped Stimulus Approach**

One of the assessments was developed by Northwest Regional Lab. It was basically a simulation of teaching situations presented through a video format. We’d have a group of teachers—new teachers—in a room watching a video presentation by a reading/language arts teacher. Then there would be questions that the viewers would respond to. The candidates are given a story to read, the same story that the children in the video would be reading, so they know what materials the teacher was using at the time. Then they answer questions about the materials and the techniques the teacher used. This was a very powerful approach, but it’s very tricky to get things just right when you are doing a video. When the teacher does something that is not quite accepted practice, people often assume you are promoting that, and we have a very strong curriculum framework in our state—any deviation from which is heresy.

There are many elements in the philosophical approach of an assessment that must be consistent and compatible with the state-recommended curriculum. Basically, this assessment was built with three types of questions for each component. The first question asked something descriptive: “What did you see in this lesson?” The second question was designed for application, to determine if the examinees knew how something worked, what its strengths and weaknesses were, and what its rationale might be. The third question explored hypothetical sorts of things, such as “What would you do next?” or “How could this teacher have done this better?” There were four lessons for this, four different videotapes that tried to cover the range of language arts/reading activities that might be anticipated. Then there
was a literature-based collaborative learning activity, which is being promoted strongly by our State Department and by other professionals in California. And finally, there was something we call the “into, through, and beyond responses” to a piece of literature.

The Assessment Center / Portfolio Approach

The second type of assessment that we piloted used the assessment center approach together with a portfolio. This was designed for secondary English teachers in our state. I was reminded of this approach yesterday during a discussion about the educative process of assessment—that an assessment is also a learning experience. Catharine Lucas, the professor from San Francisco State University who developed this approach, has very strong feelings about it. She sent out a packet of instructional materials to the candidates before they ever came into the teaching center, materials they may or may not have come across in their undergraduate preparation, such as how to respond to student writing, the nature of the writing process, so forth.

In another component of this assessment, examinees worked in a group to “construct meaning” from a literary work. We wonder if we could ever approach this degree of intensity in a large-scale assessment. There were four hours in which candidates were first given samples of student writing and asked to critique them, putting down the kind of comments they would in a real class. In addition, there was a “fishbowl” discussion of a literary work they had read in advance. They were to work in panel groups of four and interpret the story. In the “Speaking of Language” exercise, they were asked questions about how to deal with dialect differences in the classroom. Should one teach the so-called “king’s English”? These are very critical issues in teaching English.

After the four assessment center activities were completed, the candidates worked with veteran English teachers to plan a unit of instruction to be used during their first semester of actual teaching. They followed guidelines for producing portfolios documenting the unfolding of the unit. The results of these portfolios were later evaluated by experienced assessors.
The Subject-Specific Classroom Observation

Another of our assessments was of the teacher’s ability to demonstrate certain skills in an actual classroom setting. This assessment was developed by RMC and is an adaptation of the CCI that was developed in Connecticut. One of the things that came up during the validity study of the specialty area tests was that science teachers had to demonstrate an ability to conduct labs effectively and safely. In this case, the content area and the materials and equipment components were combined. But other than that, it basically used all of the same surrounding mechanisms as the CCI, which are preobservation, scripting the lesson verbatim, and then postobservation discussion. This was piloted for both elementary and secondary teachers to see how well they were able to do lab activities. Often the teachers didn’t have much in the way of materials, but sometimes implementing something like this will leverage a district to provide a better environment for science, or any other area.

Constructed-Response Assessments

One of the other assessments that we worked on, and I think there was a discussion of yesterday, used the “shell” that Steve Klein from the Rand Corporation worked with us to develop. The shells are patterns of lessons and descriptions, hypothetical descriptions, of what goes on in teaching. We piloted six of these. This is a very powerful, low-cost approach because it uses the essay format for response. So it’s a written response, which of course is much easier to deal with than some of the other formats we looked at.

Then we experimented with counterbalanced groups of beginning and experienced teachers, giving them a pretest, then some training, and then a posttest. This was developed by Northwest Regional Lab by Rick Stiggins. Many of you have probably heard of his work. He gave workshops to half of the teachers participating in this, after which we administered a posttest. He used a very pragmatic, down-to-earth approach to measurement rather than a more technical or statistical approach to measurement. We found that this is a weakness in our California requirements for teachers. They often get no help at all in assessment.
I’ll give you an example. [transparency] These were different scenarios of what might happen in the classroom. This one focuses on informal assessment that goes on when the teacher is in the process of questioning students. You notice that there are factors here related to ethnic diversity and to gender issues. When he presented the scenario, he asked some rather pointed questions about what was the appropriate assessment approach.

I think this briefly summarizes the assessments that we developed and pilot tested last year. Now we are looking at the potential of piloting interactive videodiscs in this context. We are learning a tremendous amount from piloting these assessments.