

Fresno Assessment
of Student Teachers:
A Teacher Performance Assessment
that Informs Practice

**Colleen W. Torgerson, Susan R. Macy,
Paul Beare, & David E. Tanner**
California State University, Fresno

Traditional measures of teachers' competency have been widely criticized for their lack of authenticity and predictive validity (Darling-Hammond, 2001; Porter, Youngs, & Odden, 2001). There is little evidence regarding the technical soundness of traditional teacher licensure tests and little research documenting the validity of such tests for identifying competent teachers or effective teaching (Mitchell, Robinson, Plake, & Knowles, 2001). Growing evidence indicates that performance assessments better evaluate instructional practices than these traditional assessments (Mitchell et al., 2001) and that performance assessments can serve as valuable professional learning experiences (Darling-Hammond & Snyder, 2000). Performance assessments allow for the evaluation of both the process used in solving a task and the product itself (Lane & Stone, 2006) and include evidence from actual teaching practice to potentially provide direct rather than inferred evaluation of teaching ability (Pecheone & Chung, 2006a). In 1998 California passed Senate Bill (SB) 2042 requiring teacher candidates to successfully complete a

Colleen W. Torgerson is the associate dean and a special education faculty member, Susan R. Macy is a full-time lecturer following a long career as a special educator, school site principal, and school board member, Paul Beare is the dean, and David E. Tanner is a professor in the Department of Educational Research and Administration, all with the Kremen School of Education and Human Development at California State University, Fresno, Fresno, California.

teacher performance assessment (TPA) prior to obtaining a preliminary teaching credential, and in July 2008, SB 1209 put the law into effect.

Programs in California had two options: use the TPA designed for the state by the Educational Testing Service or develop their own. To date, only two such alternative assessments have been approved for use: the Performance Assessment for California Teachers (PACT) developed by a consortium of pre-service teacher preparation programs (Chung, 2008), and the Fresno Assessment of Student Teachers (FAST), the only locally designed Commission on Teacher Credentialing (CTC) approved assessment system. This article describes the development and implementation of FAST.

The genesis of change in teacher education is often born out of either necessity or serendipitous circumstance; both were the case with FAST. California State University, Fresno (Fresno State) had nearly ten years of experience using TPAs as a means of informing the practice prior to the mandated implementation date. It was a logical next step to meet the accreditation assessment standards by utilizing faculty expertise with TPAs to develop a system that would meet state requirements and the needs of candidates and program faculty.

Teacher Work Sample (TWS) is a performance based assessment tool that enables teacher education programs to examine evidence of student teachers' ability to meet state and national teaching standards (Watkins & Bratberg, 2006; McConney, Shaylock, & Shaylock, 1998; The Renaissance Partnership for Improving Teacher Quality, 2004). Kohler, Henning, and Usma-Wilches (2008) found that TWS allowed the authors to effectively evaluate student teacher instructional decision making processes and identify relative strengths and weaknesses therein. This process allows both individual student teacher weaknesses and teaching practices to be acknowledged and remediated and to address weaknesses across the program.

As noted by Darling-Hammond and Snyder (2000), "If such [teacher performance] 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 be overwhelming within the institutional constraints of teacher education programs" (p. 527). The development of FAST was intensive with regard to time, labor, and expense but resulted in an "embedded assessment." At peak periods in its development, it was embraced with the "enthusiasm, energy, and optimism" Mehrens (1992, p. 3) associated with those doing research on performance assessment.

The Circumstances

Fresno State was one of the founding universities of The Renaissance Group (TRG), a national consortium of institutions with a commitment to the preparation of educational professionals as an “institution-wide endeavor.” TRG espouses a set of operating principles to guide its pursuit of quality and best practices in teacher education and strives to be a proactive force for the improvement and reform of education (The Renaissance Group, 2008).

Between 1999 and 2005, Fresno State was one of eleven TRG universities to participate in a multi-million dollar Title II grant for improving teacher quality. This provided money, motivation, and the collective expertise of 11 teacher education programs from across the country for faculty to spend six years developing, piloting, and refining the Teacher Work Sample (TWS). TWS is a TPA that provides evidence of a student teacher’s ability to meet state and national teaching standards while providing feedback in a form that allows for continuous program improvement (Kohler, 2008). Based on pioneering work out of the University of Oregon (Shalock & Myton, 1988), initial involvement was purely a scholarly developmental activity, not recognized as potentially useful for evaluating teacher candidates at the institutional level. Participation at Fresno State involved marked effort from university faculty, fieldwork supervisors, Beginning Teacher Support and Assessment (BTSA) partners, supervising teachers from both Multiple Subject (MS) and Single Subject (SS) programs, and advisory groups. The TWS addressed movement toward outcome measures (Cochran-Smith, 2003) and is a respected instrument that “requires the teacher candidate to systematically connect teaching and learning” (Girod & Girod, 2008, p. 309).

The impetus for the development of a local teacher performance assessment system at Fresno State was the need for assessments that informed practice and supplied data in advance of an impending National Council for the Accreditation of Teacher Education (NCATE)/CTC accreditation site visit in March 2006. This meant the system needed to be in place by spring 2003 to be fully implemented in time to generate, analyze, and report the full year of candidate performance data required by NCATE. It was not until 2003 that the CTC approved components necessary to begin the development of an instrument or procedure. Fresno State could not wait for the California TPA development and still be ready in time.

Simultaneous with the development of FAST, the Multiple Subject credential program reduced from 40 to 34 units, requiring complete revision of all the courses in that program. Through discussion and

redesign of the program's scope and sequence, the framing for the FAST was embedded across courses allowing for the logical integration of both formative and summative evaluations of candidate mastery of Teacher Performance Expectations (TPEs) at strategic points within the programs.

Development of the Assessment System

Motivated by faculty interest and timeline mandates, the efforts to develop a teacher performance assessment system began in earnest in spring 2002. Although the CTC Assessment Design Standards had not yet been established, Fresno State did use the state's short list of essential components in its own system. Operating principles included:

- all candidates would be measured against each of the TPEs at least twice;
- assessments would occur over the entire course of the teacher preparation program;
- fieldwork-based summative assessment would follow course-work-based formative assessment; and
- common performance assessments would be used in the MS and SS programs.

The goal was to measure important objectives that "cannot be easily measured by multiple choice tests" (Mehrens, 1992, p. 8). Fresno State faculty determined that TWS would be the cornerstone of this assessment system.

Research in the areas of teacher assessment, program evaluation, and performance assessment guided planning efforts. The inclusion of planning partners from a broad swath of the university in both the development of the tasks and the implementation system was strongly supported by Cochran-Smith (2006) who noted that the power to reinvent the teaching profession is an all-university responsibility, a credo which, as noted earlier, is the main uniting theme of The Renaissance Group (2008).

Authentic assessment such as Teacher Work Sample "may shape professional preparation programs in ways that encourage better integration of knowledge within and across courses and other learning experiences" (Darling-Hammond & Snyder, 2000, p. 527). The development of FAST was supported by research concerning portfolios that assemble artifacts. Such exhibitions can capture important attributes of teaching and reasoning about teaching (Darling-Hammond & Snyder, 2000). These practices may transform the teacher candidate's understanding of theory into practice.

Baratz-Snoden (1990) advised that an assessment used for performance accountability had to be professionally credible, publicly acceptable, legally defensible, and economically feasible. To meet this list of characteristics, assessment tasks were developed by committees that included content area faculty, field supervisors, Beginning Teacher Support and Assessment (BTSA) program coordinators, and supervising teachers from local districts. The TPEs needed to be taught and formatively assessed in specific coursework followed by summative assessment using a complex performance task in an authentic fieldwork context that was commensurate with the candidates' standing in the program's sequence. The tasks included some elements effective in evaluating teacher candidates in the past, including writing lesson plans, teaching the plan, developing a unit of study, and creating a teaching portfolio. Faculty supported the tasks as the complex application of knowledge and skills taught in coursework, and BTSA supported the tasks because of their close process alignment to induction level assignments.

From this initial work, teams: (1) developed specific tasks that evaluated specific TPEs; (2) designed task specific rubrics that qualitatively defined selected elements of the TPEs being evaluated; (3) field-tested the tasks with cohort groups; (4) scored performances and collected anecdotal impressions from supervising teachers, field work supervisors, and teacher candidates; (5) revised tasks and/or rubrics; and (6) field-tested, again. Following three semesters of work, the tasks were piloted in fall 2004 by all Fresno State teacher candidates. Data were collected in fall and spring, analyzed, and reported in anticipation of the March 2006 accreditation site visit. Fresno State was adjudged as meeting all assessment standards of both NCATE and CTC.

Over the next year Fresno State continued to refine rubric language and improve scorer training and calibration procedures. Calibration is the process by which an assessor's scores for a specific performance relative to a specific rubric come to match scores determined by experts to be reflective of that same performance using the same rubric. Once initially calibrated, scorers must re-calibrate annually in order to continue to score candidate performances.

In December 2006 the CTC issued its Assessment Design Standards (CTC, 2006) and provided programs with a procedure for submitting an alternative system. This required that Fresno State further refine FAST to meet the CTC's rigorous standards.

Final Product: Fresno Assessment of Student Teachers

The FAST system consists of four complex tasks administered over the span of a candidate's pre-service training that measures performances relative to the 13 TPEs. Each TPE is measured twice, using a different format in a different teaching context each time (See Figure 1). All projects are aligned with the candidate's student teaching practica. Three tasks have an accompanying rubric that generates a discreet score for each TPE evaluated by that task, the exception being the Teaching Sample Project that is scored by sections that are aligned with identified TPEs. Scores range from one to four. A score of one "doesn't meet expectations" is failing; two "meets expectations" represents passing at a competent level; three "meets expectations at a high level"; four "exceeds expectations" and has been informally described by local BTSA partners as representing the expectation for performance following the induction period.

Task directions and rubrics are provided to each candidate in the *FAST Manual* (2008) and electronically. In addition, the *FAST Manual* provides policies regarding intended use, accommodations for students with disabilities, and appeal procedures.

The four projects are the Comprehensive Lesson Plan, Site Visitation, Holistic Proficiency, and Teaching Sample Project. Figure 2 describes the tasks, when they are administered, and who scores them.

The Projects

Comprehensive Lesson Plan Project. This paper-pencil task assesses a candidate's ability to analyze a lesson plan designed for all students in a classroom (Grades 4-8) with a significant number of English learners. The analysis is evidenced through answers to specific questions provided to the candidate prior to the assessment. Sample questions are:

- What specific strategies in the lesson are used to help English Learners understand specific content information? Why do you think they are effective?
- Students in grades 4-8 are in cognitive and social transition. Describe this transition using your knowledge of Piaget, Erikson, or Vygotsky and then share the instructional activities or strategies you selected as appropriate for the students in these grades.

Site Visitation Project. This project assesses the candidate's ability to plan, implement, and reflect upon instruction. Supervisors evaluate the candidate's ability to write a lesson plan as part of on-going instruction

Figure 1

California Teaching Performance Expectations																
When Administered	Task	Making Subject Matter Comprehensible to Students			Assessing Student Learning			Engaging and Supporting Students in Learning			Planning and Designing Learning Experiences for Students		Creating and Maintaining Effective Environments for Student Learning		Developing as a Professional Educator	
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)		
When Administered	Task	(1) Specific Pedagogical Skills for Subject Matter Instruction	(2) Monitoring Student Learning During Instruction	(3) Interpretation and Use of Assessment	(4) Making Content Accessible	(5) Student Engagement	(6) Developmentally Appropriate Teaching Practices	(7) Teaching English Learners	(8) Learning About Students	(9) Instructional Planning	(10) Instructional Time	(11) Social Environment	(12) Professional, Legal, and Ethical Obligations	(13) Professional Growth		
1 SS: Initial Student Teaching	Task Lesson Plan	✓ MS only (EIA 4-8)					✓	✓	✓							
2 SS: Initial Student Teaching	Task Site Visitation	✓ MS only (EIA K-3)	✓		✓	✓							✓			
MS: Phase 3 (final Student Teaching) SS: Final Student Teaching	Task Holistic Proficiency	✓ SS, MS (R-SS, MS (Math Science))	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			
	Task Teaching Sample Project	✓ SS, MS (variable)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			



Fresno Assessment of Student Teachers
Alignment with Teaching Performance Expectations

Figure 2

FAST Tasks	Task Description	Venue/Scorers	Semester
Comprehensive Lesson Plan (TPEs 1A/6B, 7, 8, 9)	Given a prompt with the teaching context of a classroom with a significant number of EL students, student descriptions, & lesson plan, the candidate answers analysis questions	2 hour session at central site; All program faculty	Semester 1 (MS & SS)
Site Visitation (TPEs 1,2,4 5,11,13)	Candidate plans a detailed lesson (SS- content; MS – ELA), are observed teaching, & a self-evaluation of lesson	20 minute lesson taught & observed; University Supervisor/ Master Teacher	Semester 1 (SS) Semester 2 (MS)
Holistic Proficiency (TPEs 1,3,5 6,10 ,12)	Candidate documents competence through observation, artifacts provided, & self-assessment of progress on each TPE	Entire semester documentation; University Supervisor/ Master Teacher	Semester 2 (SS) Semester 3 (MS)
Teaching Sample Project (TPEs 1,2,3, 4, 7,8,9,10, 11,12,13)	Candidate plans, implements and reflects on teaching a unit of study to include: Students in Context; Content Analysis & Learning Outcomes; Assessment Plan; Design for Instruction; Instructional Decision-making; Analysis of Student Learning; Reflection	Plan & teach a 1-4 week unit; MS – All program faculty SS – Content Supervisor/ Master Teacher	Semester 2 (SS) Semester 3 (MS)

in his/her field placements, teach that lesson, and evaluate the planning and teaching of the lesson based on students' learning.

Holistic Proficiency Project. This task resembles a portfolio and

assesses the candidate's ability to perform, document, and reflect upon teaching responsibilities over an entire semester. The candidate is assessed based on direct observation of standards-based instruction, review of detailed evidence in artifacts such as student activities, pictures, student work, and self-reflections on growth and expertise for each TPE.

Teaching Sample Project. This comprehensive task is administered in final student teaching and is the cornerstone of the system, based on the TRG Work Sample. This project assesses the candidate's ability to plan and teach a 1- to 4-week unit, to assess students' learning related to the unit, to document students' learning, and to reflect on their own teaching. Specific directions and rubrics are provided for the seven sections:

- **Students in Context:** identify characteristics and factors for instructional design, including classroom management;
- **Content Analysis and Learning Outcomes:** select content standards and develop learning outcomes;
- **Assessment Plan:** adapt or develop assessments to plan, monitor, and measure student progress of learning outcomes;
- **Design for Instruction:** design overview of unit and lesson plans based on pre-assessment results;
- **Instructional Decision-Making:** provide two examples of instructional decision-making based on students' learning or responses;
- **Analysis of Student Learning:** analyze assessment data and represent data from whole class and subgroups in visual and narrative forms; and
- **Reflection and Self-Evaluation:** reflect on performance, make suggestions for improvement, and identify future goals for professional growth.

The FAST product was designed to require candidates to continually connect theory to practice and to grow instructionally across each semester of the program.

Figure 3 represents examples of the assessment of TPE 7 (Teaching English Learners) across the FAST tasks that provides a picture of the sequential and growing knowledge in the area of English Learners.

Figure 3

<i>TPE - Description</i>	<i>Task</i>	<i>Rubric Descriptor (level 2)</i>
7 – Teaching English Learners <ul style="list-style-type: none"> • Using students’ assessed levels of English proficiency; • Differentiated instruction; • Making content accessible to students 	Comprehensive Lesson Plan Project (CLPP): Candidates identify strategies within the lesson that make content accessible to English learners of various levels of English proficiency relative to the English language levels of English learners described in “Students and the Teaching Context” section	“Candidate accurately describes at least two general instructional practices in the lesson plan used to help English Learners in the class understand the content and provides a general rationale for their effectiveness...and recommends an additional or alternative strategy...”
7 – Teaching English Learners <ul style="list-style-type: none"> • Differentiated instruction; • Making content accessible to students; • Systematic instruction 	Teaching Sample Project: Assessment Plan—candidates are asked to specify assessment adaptations for English Learners . . .	“Some assessment adaptations for EL . . . students are generally appropriate.”
7 – Teaching English Learners <ul style="list-style-type: none"> • Differentiated instruction; • Making content accessible • Systematic instruction 	Teaching Sample Project: Design of Instruction—candidate is required to describe how three lessons were or could be adapted for English learners.	“Some ideas for differentiating instruction are described, including instruction of English learners . . .”
3 – Interpretation and Use of Assessment <ul style="list-style-type: none"> • Accurately interpret test results 	Teaching Sample Project: Analysis of Student Learning—Candidates are asked to identify evaluate the learning of English learners by comparing this subgroup’s learning to that of the rest of the class.	“Includes some evidence of the impact on student learning related to the learning outcome. Beginning to accept responsibility for the success of all students.”

Figure 3 (continued)

<i>TPE - Description</i>	<i>Task</i>	<i>Rubric Descriptor (level 2)</i>
3 – Interpretation and Use of Assessment <ul style="list-style-type: none"> Identify proficiency of English learners 	Teaching Sample Project: Students in Context— Candidates are asked to identify levels of English learners and the implications for instruction	“Factors selected are generally relevant to instruction. Description of implications appropriate to instruction in general.”
12 – Professional, Legal, and Ethical obligations <ul style="list-style-type: none"> Access to opportunities to learn content; Awareness of personal values and biases. 	Teaching Sample Project: Reflection and Self-Evaluation— Candidates reflect upon the implications of personal biases and how they did or will in the future ensure that English learners had appropriate opportunities to learn the content of their unit.	“Identifies successful activities or assessments and explores reasons for their success (no use of theory or research). Suggests some instructional techniques for English learners.... Evidence of seeing some connections between learning outcomes, instruction, assessment, or subject matter knowledge.”
12 – Professional, Legal, and Ethical Obligations <ul style="list-style-type: none"> Implications of policies and procedures related to English learners 	Holistic Proficiency Project - Candidates are required to reflect upon their awareness of policies and procedures related to English learners	“...Reflection shows an awareness of the implications of district, state or federal policies and procedures pertaining to the education of English learners.”

Reliability and Validity

The usefulness of performance assessment for licensure and program improvement depends on the degree to which the scoring is valid and reliable. Evaluating the validity of FAST for its ability to accurately and fairly measure the teaching skills of teacher candidates is critical. However, scores can be no more valid than they are reliable; reliability

coefficients represent a ceiling to validity measures (Huck, 2008). Linn, Baker, and Dunbar (1991) cautioned that it would be unreasonable to assume that group differences that are exhibited in traditional assessment would be alleviated by using performance assessment. This underscores the need to demonstrate the assessment system's fairness to gender and ethnic groups (Lane & Stone, 2006).

Reliability

The nature of performance assessments introduces a level of complexity in achieving inter-rater reliability unknown in more traditional testing. Jones, Jones, and Hargrove (2003) stated simply, "Portfolios and other types of authentic assessments have greater subjectivity in the scoring process and as a result, tend to have lower reliability" [than more conventional assessments] (p. 50). Schafer, Gagne, and Lissitz (2005) chronicled many of the reasons why expecting anything like the reliability associated with multiple-choice type assessments is unreasonable in a performance assessment. As may be noted, "The Teacher's Guide for the Writing" supplement of the *Iowa Test of Basic Skills* reports inter-rater reliability scores of .48 for essays using the same mode of discourse (Hieronymus, Hoover, Cantor, & Oberley, 1987, p. 28). Dunbar, Koretz and Hoover (1991) reported inter-rater reliabilities for a number of performance assessment studies with values from .26 to .60. In what may have been a premature obituary, Parkes (2007) noted, "The performance assessment movement of the 1980s and 1990s waned largely because large scale performance assessment scores struggled to, but never did achieve sufficient reliability" (p. 2).

By way of contrast, those involved with FAST have worked diligently to accomplish what Parkes (2007) noted has generally been out of the reach of proponents of performance assessment. As may be seen in Table 1, in 54% of the 248 possible decisions on the Holistic Proficiency, the first scorer and the second scorer were in absolute agreement in the January norming task. There was 71% agreement in May. The probability that this could have occurred by chance is slightly less than $p = .25$. In none of the instances was there disagreement about whether the student passed the project, or to put it in the positive, as a measure of whether the student passed or failed the Holistic Proficiency Project, agreement was 100%. Of the 113 disagreements, on that task 1% were 2 or more points apart. None of the disagreements concerned whether the student passed the particular task. Overall exact match scoring was 69.76% in January, 71.71% in May.

Regarding the Teaching Sample Project, scorers did not disagree over whether a student successfully completed the project as a whole,

Table 1
Summary of Numbers and Overall Percentages of Exact Matches
and Disagreements for the Four FAST Tasks

<i>January Administration</i>					
The Task	Total Possible Decisions	Exact Match	+/-1 Point	+/-2 more	Pass/Fail Disagreements
Comprehensive Lesson Plan	165	129	31	5	12
Teaching Sample	217	129	79	9	16
Site Visitation	210	193	17	0	0
Holistic Proficiency	248	135	110	3	0
Percent		69.76%	28.21%	2.02%	3.33%
<i>May Administration</i>					
The Task	Total Possible Decisions	Exact Match	+/-1 Point	+/-2 more	Pass/Fail Disagreements
Comprehensive Lesson Plan	110	79	31	0	7
Teaching Sample	66	53	11	2	0
Site Visitation	182	125	55	2	0
Holistic Proficiency	144	103	40	1	1
Percent		71.71%	27.29%	<1%	1.59%

but only whether the student passed a particular component of the task. With seven different components scored, these disagreements generally regarded one of the seven components within the project. Obviously, had FAST been scored 'holistically,' with the entire project pass/fail, a higher reliability could have been obtained. The impact of a disagreement over a single component of the entire task is ameliorated by the fact that students who receive a failing grade can remediate and resubmit. In this regard, the project is unlike a traditional high stakes assessment.

By any published standard for performance assessment identified, the level of inter-rater reliability that was achieved here is higher than the norm. The most similar instrument identified for comparison was the PACT. An examination of data from the *Technical Report for PACT* (Pecheone & Chung, 2006b) showed a 56.57% exact match as compared to FAST's 69.76% figure.

Validity

Among other criteria, the validity of FAST refers to the appropriateness, meaningfulness, and utility of the candidate produced work that is used to support decisions on a candidate's recommendation for an initial teaching credential. Demonstrated validity also allows for faculty members to improve the quality of the credential program. FAST was specifically created to meet the requirements of SB 2042 that a candidate show "proficiency" on the TPEs prior to being recommended for a state licensure. When dealing with performance assessment, the language may differ from that associated with traditional assessment (Lane & Stone, 2006). Rather than referring to construct, or criterion-related validity, for example, in addition to the reliability described above, Fredriksen and Collins (1989) proposed examining "directness, scope, ..., and transparency" (p. 30) as criteria for the validity of a performance assessment.

Directness refers to explicitly assessing the desired knowledge and skills. FAST content explicitly represents the 13 TPEs identified by the CTC. In a term analogous to content validity, *scope* refers to covering all the knowledge, skills, and strategies required to do well in an activity, in this case, teaching. FAST covers the entirety of the California TPEs which were established by policy makers, teachers, teacher educators, and administrators based on a statewide job analysis (Pechione & Chung, 2006). A panel of expert teacher educators and teachers participated in the development of the tasks associated with each TPE to verify the content was an authentic representation of an important dimension of teaching. Scope and directness together are a form of content validity. *Transparency* is the degree to which the terms of judgment are clear to those taking an assessment. Fredriksen and Collins (1989) argued that instruments must be transparent enough so that those taking it can assess themselves and others with almost the same accuracy as the actual evaluators. The rubrics for scoring all the TPEs for each of the tasks are provided to teacher candidates and reviewed for them repeatedly in the course of their program. *Reliability* was described above. Clearly, FAST meets the validity standards for performance assessment identified by Fredriksen and Collins (1989) and described by Lane and Stone (2006).

Gender and Ethnicity Fairness

Basic analyses were completed to identify any differential effects in relation to candidates' ethnic group or gender on FAST's four tasks. The Kruskal-Wallis H, a non-parametric test for significant differences among more than two groups when the dependant variable is an ordinal scale, was used to assess ethnic differences. For gender differences, the Mann-Whitney U was performed for each TPE on all four FAST tasks

for both MS and SS candidates. The only significant difference ($p=.05$) on ethnicity was for TPE 9 on the Comprehensive Lesson Plan where Hispanic candidates scored lower than other groups. There were differences by gender on three of the TPEs on the Comprehensive Lesson Plan with males scoring lower; however, this finding should be tempered by the small size of the male group (10% of the whole). The only other significant difference by gender was on Site Visitation where females scored higher on TPE 5. None of these differences was great enough to affect overall passing scores on any task. In each instance there is on-going review by faculty to determine that differences do not stem from insensitivity to the lower scoring group.

Selecting, Training, and Calibrating Assessors

All FAST projects are scored by trained assessors coming from the faculty in teacher education or single subject content areas, master teachers, student teaching supervisors, and local BTSA support providers. Each assessor is trained, periodically tested, and must meet calibration standards annually in order to score candidate performances. A database is maintained to identify qualified scorers who meet the FAST criteria: pedagogical expertise, completed project-specific training, and calibration on the project(s) within one year of scoring the task.

The basic design of each task's scorer training is the same and includes the following elements: assessor guidelines, bias training, and calibration and re-calibration of scorers. Scorers are given a copy of the project directions, the corresponding project-specific rubric, and provided with an overview of the project by the project trainer. After scorers familiarize themselves with the expectations of students' performances, the trainer presents critical guidelines that should guide scoring: rely on the rubric as the sole criteria for scoring each performance; maintain an attitude of respect for all performances; understand that excellent teaching takes many forms; do not be fooled by writing ability or other elements not evaluated by the project; and avoid the common pitfall of scoring such as the inference of a positive (or negative) performance on one section based on performance on another part of the task. Scorers then complete an activity that involves discussing biases and how those biases about excellent or poor teaching can influence their evaluation of candidate responses.

Scorers are organized in pairs or trios of experienced and inexperienced scorers to find a common understanding of the rubric by highlighting strategic words or phrases that qualitatively differentiate one level of the rubric from another. Experienced scorers independently score a marker performance, comparing their own score for each TPE against

scores already established by a team of experts. If the scores conform, the scorer is considered re-calibrated and is authorized to score the task. Inexperienced scorers, however, still work in pairs or trios to collaboratively score a second marker performance. The scores and rationale used to determine the score are shared with the entire group. Based on predetermined scores and a written rationale, the trainer clarifies any misconceptions. An inexperienced scorer scores a third marker performance independently, comparing scores for each TPE evaluated against scores established by a team of experts. If the scores conform, the scorer is considered calibrated and is authorized to score the task. Experienced and inexperienced scorers whose scores fail to align with those awarded by experts will score with a calibrated scorer until their scores fall into alignment at which time they are allowed to score independently. Uniformity in scorer training enhances reliability and validity and provides for input from an array of expert scorers with multiple pedagogical perspectives. Such detailed scoring protocols increase time spent on scorer training but have been shown to dramatically reduce errors in measurement due to unreliable raters (Dunbar, Koretz, & Hoover, 1991).

Securing CTC Approval

The CTC Assessment Design Standards (CTC, 2006) required that TPAs be valid, fair, and at least as rigorous as the state passing standards. The issuance of specific standards was helpful and stimulated new conversations within and between programs to come to agreement as to the formal policies and procedures that would govern and 'systemize' the system. To this end, Fresno State submitted a 160-page written document to the Commission in early June 2007 that addressed the eight elements aligned with each standard and included two appendices: *FAST Tasks and Rubrics* and *Scorer Training Procedures*. The CTC's Assessment Review Team responded to Fresno State's submission within the month by approving six of the elements outright, approving parts of six other elements, and requesting more information with regard to the remaining six elements. The specificity of the Review Team's critique of the tasks and scoring rubrics was extremely helpful and instigated changes that strengthened the assessment system.

In October, 2007, Fresno State submitted revised FAST tasks and rubrics, as well as data charts on which analyses were founded. Within a month, the Assessment Review Team acknowledged the clarification of task and rubric statements and requested reliability data generated from the revised assessment tools and rubrics. These data and their analysis were provided for fall 2007 and spring 2008. Finally in May

2008, the Assessment Review Team recommended that the FAST model be approved by the Commission, and in an action at their June 5, 2008 CTC meeting, the FAST was approved as an alternative TPA model.

Using the Data

Teacher candidates are informed by their fieldwork supervisor as to his or her level of performance on specific FAST tasks. A candidate who earns a score of “one, does not meet expectations,” is provided with remedial instruction and is given the opportunity to attempt the failed task again. All scores earned are tracked and used locally for statistical purposes only. Only passing scores are included in the candidate report, in that candidates who fail are not recommended for the credential.

Informing the Program

Annually the frequency of all scores, the mode, and the median are calculated and analyzed school-wide as well as by sub-groups. The data are used by faculty for program improvement. In addition, tasks are subjected to an in-depth review and analysis every two years on a rotating basis; thus, one task is reviewed each semester, and every task will have been evaluated every two years.

A minimum of 15% of responses to each task are double-scored to determine inter-rater reliability for each TPE for each task. These data are used to evaluate scorer training and calibration. Data generated by tasks under review are analyzed by gender, ethnicity, and self-reported English language proficiency. This periodic review helps assure that FAST maintains its high level of reliability and its usefulness in informing stakeholders as to candidate and program performance.

Using FAST Data for Program Improvement:

An Early Example

Fresno State is working closely with the California State University Center for Teacher Quality (CTQ) in aligning responses to the CTQ’s annual surveys of graduates and supervisors to TPEs. An analysis at such a discreet level, over time, and from multiple sources, will provide robust data for program evaluation and improvement. Such an analysis has already occurred using FAST data with informal references to the Center’s surveys and programmatic changes implemented.

Using data generated during 2005-2006 field-testing, Fresno State found that candidates performed at a minimal level with regard to TPE 7, Teaching English Learners. Graduates with one year of professional experience and their site supervisors, as well as candidates completing

the exit survey, also reported only minimally acceptable preparation in teaching English Language Learners (ELL).

As a result, in 2006-2007 Fresno State implemented several improvement efforts related to skills in teaching ELL. Faculty meetings using recommended readings and presentations by recognized ELL experts were held. A series of seminars for faculty were presented to enhance professional knowledge and skills related to strategies such as the use of contextual clues, multi-sensory experiences, scaffolding instruction, comprehensible input, and comprehension checks. Instructional methods courses were directed to include more overt emphasis on modeled ELL teacher behaviors. An all-day retreat to a 100% ELL school district was held by the faculty to observe the strategies used and to interact with teachers, students, administrators, and parents relative to that district's ELL strategies. By tracking the FAST task scores on TPE 7 for specific groups of teacher candidates as they moved through the program, mean scores were raised from 2.32 in fall 2006 (semester 1) to 3.42 in fall 2007 (semester 3). This documentation of improved candidate knowledge and practice in teaching ELL was the desired outcome of the described activities and changes made in the credential programs.

Improving candidates' professional skills in teaching ELL is an ongoing goal, and this type of documentation allows much quicker examination of intervention effects than waiting two or three years for follow-up survey results. The alignment with survey data from the CTQ will assist and inform these efforts.

Conclusion

In contrast to other university programs that had to select a performance assessment and secure faculty support and buy-in, the Fresno State faculty effort, expertise, and investment in the creation of FAST made its adoption a natural part of a multi-year process to improve programs and assessment. Fresno State began using the Teacher Work Sample independent of California mandates and would continue to utilize TPAs if the mandate were eliminated. The knowledge gained from FAST informs practice, quickly reflects changes in program or course requirements, improves candidate skills, and ultimately improves the learning of K-12 children.

The development and use of FAST has modeled to faculty university-wide that the assessment of teaching goes beyond simply measuring one's knowledge of content. Performance assessment is a measure of the complex pedagogical skills required for candidates to successfully teach and cause their students to learn. This critical feature has served

as a model for outcomes assessment at the university. The cross-campus participation and support for FAST development would not have been possible without the President, Provost, and faculty's strong belief in The Renaissance Group ethic that teacher preparation is an endeavor that must involve and be supported by an entire campus. Participation in TRG would be judged as invaluable for this campus for that reason alone, independent of the experience with TWS that it provided.

FAST meets the criteria for an assessment system set forth by the National Board of Professional Teaching Standards as stated by Baratz-Snowden (1990). It is feasible, professionally credible, publicly acceptable, legally defensible, and economically affordable. It is with great excitement that Fresno State looks forward to both quantitative and qualitative examinations of its effects on program, candidate, and K-12 student performance.

References

- Baratz-Snowden, J. (1990). Research news and Comments: The NBPTS begins its research and development program. *Educational Researcher*, 19(6), 19-24.
- California Commission on Teacher Credentialing. (2006). Assessment design standards. Retrieved October 15, 2008 from www.ctc.ca.gov/educator-prep/TPA-files/TPA-Assessment-Design-Standards.doc.
- Chung, R. R. (2008). Beyond assessment: Performance assessments in teacher education. *Teacher Education Quarterly*, 35, 7-28.
- Cochran-Smith, M. (2003). Assessing assessment in teacher education. *Journal of Teacher Education*, 54, 187-191.
- Cochran-Smith, M. (2006). Ten promising trends (and three big worries). *Educational Leadership*, 63(6), 20-25.
- Darling-Hammond, L. (2001). Standard setting in teaching: Changes in licensing, certification, and assessment. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed., pp. 751-776). Washington, DC: American Educational Research Association.
- Darling-Hammond, L., & Snyder, J. (2000). Authentic assessment of teaching in context. *Teaching and Teacher Education*, 16, 523-545.
- Dunbar, S., Koretz, D., & Hoover, H. (1991). Quality control in the development and use of performance assessments. *Applied Measurement in Education*, 41, 289-303.
- Frederiksen, J. R., & Collins, A. (1989). A systems approach to educational testing. *Educational Researcher*, 18(9), 27-32.
- Fresno Assessment of Student Teachers Manual*. (2008). Fresno, CA: Author.
- Girod, M., & Girod, G. (2008). Simulation and the need for practice in teacher preparation. *Journal of Technology and Teacher Education*, 18, 307-337.
- Hieronimus, A. N., Hoover, H. D., Cantor, N. K. & Oberley, K. R. (1987). *Writing supplement teacher's guide: Iowa Tests of Basic Skills*. Chicago: The Riverside Publishing Company.
- Huck, S. W. (2008). *Reading statistics and research* (5th ed.). Boston: Pearson.

- Jones, M. G., Jones, B. D., & Hargrove, T. Y. (2003). *The unintended consequences of high-stakes testing*. Boulder, CO: Rowman & Littlefield Publishers.
- Kohler, F. (2008). Preparing preservice teachers to make instructional decisions: An examination of data from the teacher work sample. *Teaching and Teacher Education, 24*, 2108-2117.
- Kohler, F., Henning, J., & Usma-Wilches, J. (2008). Preparing preservice teachers to make instructional decisions: An examination of data from the teacher work sample. *Teaching and Teacher Education, 24*, 2108-2117.
- Lane, S., & Stone, C. A. Performance assessment. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 387-431).
- Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Complex performance assessment: Expectations and validation criteria. *Educational Researcher, 20*(8), 15-21.
- McConney, A., Shalock, M. D., & Shalock, H. D. (1998). Focusing improvement and quality assurance: Work samples as authentic performance measures of prospective teachers' effectiveness. *Journal of Personnel Evaluation in Education, 11*, 343-363.
- Mehrens, W. (1992). Using performance assessment for accountability purposes. *Educational Measurement: Issues and Practice, 11*, 3-9, 20.
- Mitchell, K. J., Robinson, D. Z., Plake, B. S., & Knowles, K. T. (2001). *Testing teacher candidates: The role of licensure tests in improving teacher quality*. Washington, DC: National Academy Press.
- Parke, J. (2007). Reliability as argument. *Educational Measurement: Issues and Practice, 26*(4), 2-10.
- Pecheone, R., & Chung, R. (2006). Evidence in teacher education: The Performance Assessment for California Teachers (PACT). *Journal of Teacher Education, 57*(1), 22-36.
- Pecheone, R. L., & Chung, R. R. (2006b). Technical report of the Performance Assessment for California Teachers (PACT): Summary of validity and reliability studies for the 2003-4 pilot year. Palo Alto, CA: PACT Consortium.
- Porter, A., Youngs, P., & Odden, A. (2001). Advances in teacher assessments and uses. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed., pp. 259-297). Washington, DC: American Educational Research Association.
- Schafer, W. D., Gagne, P., & Lissitz, R. W. (2005). Resistance to confounding style and content in scoring constructed-response items. *Educational Measurement Issues and Practice, 24*(2), 22-28.
- Shalock, H. D., & Myton, D. V. (1988). A new paradigm for teacher licensure: Oregon's demand for evidence of success in fostering learning. *Journal of Teacher Education, 39*(6), 8-16.
- The Renaissance Group. (2008, July 28). The Renaissance Group's purpose. Retrieved October 26, 2008, from <http://education.csufresno.edu/rengroup/>
- The Renaissance Partnership for Improving Teacher Quality. (2004). Renaissance teacher work samples. Retrieved January 5, 2009 from <http://www.uni.edu/itq/RTWS/index.htm>.
- Watkins, P., & Bratberg, W. (2006). Teacher work sample methodology: Assessment and design compatibility with fine arts instruction. *National Forum of Teacher Education Journal, 17*, 1-10.