

Predictions and Performance on the PACT Teaching Event: Case Studies of High and Low Performers

By Judith Haymore Sandholtz

Performance assessments are becoming an increasingly common strategy for evaluating the competency of pre-service teachers. In connection with standards-based reform and a focus on teacher quality, many states have moved away from relying on traditional tests or university supervisors' observations and have developed performance assessments as part of licensing requirements or accreditation of teacher education programs (Pecheone, Pigg, Chung, & Souviney, 2005). In California, legislation requires that teacher certification programs implement a performance assessment to evaluate candidates' mastery of specified teaching performance expectations (California Commission on Teacher Credentialing, 2006). The concerns about traditional tests used in licensing decisions center on the extent to which the tests are authentic and valid in identifying effective teaching (Mitchell, Robinson, Plake, & Knowles, 2001). Relying on university supervisors' classroom observations of candidates for summative judgments is also problematic due to issues of validity and reliability. A potential advantage of university supervisor assessments is that judgments are based on observation of candidates' actual teaching in classroom settings. However, observations may be conducted too infrequently, training of supervisors may be insufficient to achieve inter-rater agreement, and observation forms may not be tailored to specific disciplines or levels (Arends, 2006b). Researchers

■
*Judith Haymore
Sandholtz is a
professor in the School
of Education at the
University of California,
Irvine.*

Predictions and Performance on the PACT Teaching Event

who investigated how teacher candidates were evaluated in student teaching across multiple types of teacher preparation institutions in the U.S. reported that summative judgments made from student teaching observation forms were unable to differentiate among various levels of effectiveness (Arends, 2006a).

When performance assessments include evidence from teaching practice, they can provide more direct evaluation of teaching ability than pencil-and-paper licensure tests or completion of coursework (Mitchell et al., 2001; Pecheone & Chung, 2006; Porter, Youngs, & Odden, 2001). But, as with traditional tests and supervisor observations, concerns about the reliability and predictive validity of performance assessments must be resolved (Pecheone & Chung, 2006). Other concerns about performance assessments center on the effects on curriculum and the richness of teacher education programs, potential harm to relationships essential for learning, competing demands, and the significant amount of human and financial resources required (Arends, 2006a; Delandshere & Arens, 2001; Snyder, 2009; Zeichner, 2003). A particularly pressing issue is the high cost of developing and implementing performance assessments during periods of funding shortages (Guaglianone, Payne, Kinsey, & Chiero, 2009; Porter, Youngs, & Odden, 2001). The ongoing costs, in terms of financial support and faculty time, lead teacher educators to question if resources could be better spent in other ways (Snyder, 2009). If performance assessments provide little information beyond what university supervisors gain through formative evaluations and classroom observations of candidates, then the high costs, in combination with other concerns, may seem less justifiable.

In an earlier study, my co-author and I explored the extent to which supervisors' perspectives about candidates' performance corresponded with outcomes from a summative performance assessment (Sandholtz & Shea, 2012). We specifically examined the relationship between university supervisors' predictions and candidates' performance on the Performance Assessment for California Teachers (PACT) teaching event. We found that university supervisors' predictions of their candidates' performance did not closely match the PACT scores and that inaccurate predictions were split between over- and under-predictions. In addition, supervisors did not provide more accurate predictions for high and low performers than other candidates. Common wisdom suggests that university supervisors, who observe and evaluate candidates in the classroom, would be well-positioned to predict which pre-service teachers would perform particularly well or poorly on a teaching performance assessment. Yet, for the majority of the high and low performers, a group of 43 out of 337 candidates, their supervisors did not identify them as the exceptional candidates who would either excel or fail. For only two of 22 high performers did supervisors similarly predict high performance, and for only four of 21 low performers did supervisors predict low performance (see Sandholtz & Shea, 2012, for complete findings).

In this follow-up study, I examine the high performers and the low performers with the greatest differences between their supervisor's predictions and their PACT

scores. The aim in this study is to begin to understand how and why these differences occurred and to consider implications for assessments of pre-service candidates.

Assessment of Teaching Practice

The theoretical framework for this study draws from research establishing the relationship between conceptions of teaching and measures of effective teaching. Methods for assessing teaching practice have changed over time as conceptions of teaching have changed. During the period of process-product research, teaching effectiveness was “attributable to combinations of discrete and observable teaching performances per se, operating independent of time and place” (Shulman, 1986, p. 10). Researchers tended to control for context variables such as subject matter, age and gender of students, type of school, and ability level of students. The research frequently identified teacher behaviors that were correlated with student outcomes. These observable behaviors became the key elements of teaching effectiveness and often led to specific mandates and teaching policies for improving student achievement.

Over time, conceptions of effective teaching shifted and recognized the complex, changing situations that teachers encounter (Darling-Hammond & Sykes, 1999; National Board for Professional Teaching Standards [NBPTS], 1999; Richardson & Placier, 2001). Researchers noted the importance not only of the particular classroom context but also the larger contexts in which the class is embedded (Shulman, 1986). Teaching came to be viewed as a highly complex task that occurred in real-time, involved social and intellectual interactions, and was shaped by the students in the class (Leinhardt, 2001). In order to meet the diverse and changing needs of students in their classrooms, teachers needed to use their professional judgment and adapt their teaching. Their instructional decisions depended not only on the particular students being taught but also the specific content (Shulman, 1987). Rather than applying standard solutions, teachers needed the ability to address unique, often problematic, situations (NBPTS, 1999). The variation in context, combined with the complexity of teaching, shifted the view of the teacher to “a thinking, decision-making, reflective, and autonomous professional” (Richardson & Placier, 2001).

As conceptions of effective teaching changed, researchers noted problems with traditional measures that corresponded with a view of teaching as highly prescribed and rule-governed (Porter, Youngs & Odden, 2001; Tellez, 1996). They identified a need for assessments that better aligned with progressive, professional teaching practices. Since professionals exercise judgment and discernment in their work, assessments of teachers needed to extend beyond traditional stereotypes of classroom teaching and be more complex, open-ended, and specific to subject matters and grade levels (Haertel, 1991). To align with a view of teachers as professionals, assessments needed to recognize that teachers plan, conduct, analyze, and adapt their practices (Darling-Hammond, 1986). In particular, assessments of a candidate’s ability to teach needed to sample actual knowledge, skills, and dispositions as they

Predictions and Performance on the PACT Teaching Event

are used in teaching and learning contexts (Darling-Hammond & Snyder, 2000). In the systems of teacher assessment developed by professional organizations such as the Educational Testing Service (ETS), the Interstate New Teacher Assessment and Support Consortium (INTASC), and the National Board for Professional Teaching Standards (NBPTS), the ability to learn from one's practice is considered a central component of effective teaching. Each of their assessment systems emphasizes reflective practice and features performance-based assessments.

An advantage of performance-based assessments is the use of evidence from teaching practice (Mitchell et al., 2001; Pecheone & Chung, 2006; Porter, Youngs, & Odden, 2001). Performance-based assessments may include, for example, lesson plans, curricular materials, teaching artifacts, student work samples, video clips of teaching, narrative reflections, or self-analysis. By using evidence that comes directly from actual teaching, performance assessments offer a view into how a teacher's knowledge and skills are being used in the classroom. In addition, the documents potentially reveal how teachers reflect on their practice and adapt their instructional strategies in order to be more effective. In keeping with a professional view of the teacher, performance-based assessments offer a method for evaluating teaching ability within specific classroom contexts while also providing a potential learning opportunity for teacher candidates (Bunch, Aguirre, & Tellez, 2009; Darling-Hammond & Snyder, 2000; Okhremtchouk, Seiki, Gilliland, Atch, Wallace & Kato, 2009).

Performance Assessment for California Teachers

PACT is one of several teaching performance assessment models approved by the California Commission on Teacher Credentialing. Developed by a consortium of universities, the PACT assessment is modeled after the portfolio assessments of the Connecticut State Department of Education, the Interstate New Teacher Assessment and Support Consortium, and the National Board for Professional Teaching Standards. The assessment includes the use of artifacts from teaching and written commentaries in which the candidates describe their teaching context, analyze their classroom work, and explain the rationale for their actions. The PACT assessments focus on candidates' use of subject-specific pedagogy to promote student learning.

The PACT program includes two key components: (1) a formative evaluation based on embedded signature assessments that are developed by local teacher education programs, and (2) a summative assessment based on a capstone teaching event. The teaching event involves subject-specific assessments of a candidate's competency in five areas or categories: planning, instruction, assessment, reflection, and academic language. Candidates plan and teach an instructional unit, or part of a unit, that is videotaped. Using the video, student work samples, and related artifacts for documentation, the candidates analyze their teaching and their students' learning. Following analytic prompts, the candidates describe and justify their decisions by explaining their reasoning and providing evidence to support their conclusions. The

prompts help candidates consider how student learning is developed through instruction and how analysis of student learning informs teaching decisions both during the act of teaching and upon reflection. The capstone teaching event is designed not only to measure but also to promote candidates' abilities to integrate their knowledge of content, students, and instructional context in making instructional decisions and to stimulate teacher reflection on practice (Pecheone & Chung, 2006).

The teaching events and the scoring rubrics align with the state's teaching standards for pre-service teachers. The content-specific rubrics are organized according to two or three guiding questions under the five categories identified above. For example, the guiding questions for the planning category in elementary mathematics include: How do the plans support students' development of conceptual understanding, computational/procedural fluency, and mathematical reasoning skills? How do the plans make the curriculum accessible to the students in the class? What opportunities do students have to demonstrate their understanding of the standards/objectives? For each guiding question, the rubric includes descriptions of performance for each of four levels. According to the implementation handbook (PACT Consortium, 2009), Level 1, the lowest level, is defined as not meeting performance standards. These candidates have some skill but need additional student teaching before they would be ready to be in charge of a classroom. Level 2 is considered an acceptable level of performance on the standards. These candidates are judged to have adequate knowledge and skills with the expectation that they will improve with more support and experience. Level 3 is defined as an advanced level of performance on the standards relative to most beginners. Candidates at this level are judged to have a solid foundation of knowledge and skills. Level 4 is considered an outstanding and rare level of performance for a beginning teacher and is reserved for stellar candidates. This level offers candidates a sense of what they should be aiming for as they continue to develop as teachers.

To prepare to assess the teaching events, scorers complete a two-day training in which they learn how to apply the scoring rubrics. These sessions are conducted by Lead Trainers. Teacher education programs send an individual to be trained by PACT as a Lead Trainer or institutions might collaborate to develop a number of Lead Trainers. The training emphasizes what is used as sources of evidence, how to match evidence to the rubric level descriptors, and the distinctions between the four levels. Scorers are instructed to assign a score based on a preponderance of evidence at a particular level. In addition to the rubric descriptions, the consortium developed a document that assists trainers and scorers in understanding the distinctions between levels. The document provides an expanded description for scoring levels for each guiding question and describes differences between adjacent score levels and the related evidence. Scorers must meet a calibration standard each year before they are allowed to score.

Predictions and Performance on the PACT Teaching Event

Methods

The initial study included 337 candidates enrolled in a public university's teacher education program over two years (see Sandholtz & Shea, 2012). In that group, there were 22 high performers with a total score of 37 or higher (out of a possible total score of 44) and 21 low performers with a total score of 20 or less. The cut-off scores of 37 and 20 fell at the end of the second standard deviation of the total scores and meant that the candidates received a ranking on at least one question that was at the low or high end of the rubric scale. For this follow-up study, I identified the four high performers and the four low performers with the largest differences between predictions and total scores on the PACT teaching event. Since classroom observation data were missing for two of the four high performers, I replaced them with the two high performers with the next largest difference between predictions and scores. For each candidate, both the predictions and scores included a ranking from 1 to 4 on each of 11 guiding questions grouped in five categories. Table 1 summarizes the focus of the guiding questions at the time of data collection. As described above, the rankings are defined as: Level 1—not meeting performance standards; Level 2—acceptable level of performance; Level 3—advanced level of performance relative to most beginners; Level 4—outstanding and rare level of performance for a beginning teacher (PACT Consortium, 2009).

In this program, all of the university supervisors also acted as scorers for the performance assessment, though typically not for their own advisees. In keeping with the training outlined by the PACT Consortium, the supervisors at this university participated in two days of training each year. A Lead Trainer, who works in the

Table 1
Focus of Guiding Questions in PACT Rubrics*

Category	Focus of Guiding Questions
Planning (Questions 1,2,3)	Q1: Establishing a balanced instructional focus Q2: Making content accessible Q3: Designing assessments
Instruction (Questions 4, 5)	Q4: Engaging students in learning Q5: Monitoring student learning during instruction
Assessment (Questions 6, 7)	Q6: Analyzing student work from an assessment Q7 Using assessment to inform teaching
Reflection (Questions 8, 9)	Q8: Monitoring student progress Q9: Reflecting on learning
Academic Language (Questions 10,11)	Q10: Understanding language demands Q11: Supporting academic language

*Note: PACT=Performance Assessment for California Teachers. An additional question on assessment as added in 2009-10.

teacher education program and had been trained by PACT for the role, conducted the sessions. During the training, the supervisors scored two or three benchmark teaching events, which were provided by the PACT Consortium. Each person read a specific section of the event, assigned a score, and compared their scores with the benchmark scores. The group then discussed any variations. Following the training, and before being allowed to score teaching events, the supervisors had to pass a calibration standard set by the PACT Consortium. Each supervisor's scores on the calibration teaching event (provided each year by PACT) had to meet three criteria: a) resulted in the same pass/fail decision; b) included at least six exact matches out of the 11 rubric scores; c) did not include any scores that were two away from the pre-determined score. After completing training for PACT scoring and passing the calibration standards, university supervisors predicted scores for their own candidates and then received their assigned assessments to score. The training, calibrating, predicting, and scoring took place within a two-week period.

As trained scorers, the supervisors understood the PACT and the scoring levels, but they were not directly involved in preparing their candidates for the performance assessment. In this program, the supervisors did not teach courses or seminars for student teachers. The supervisors' role was to provide support and guidance for student teachers in their assigned classrooms. Over the academic year, supervisors made ongoing, periodic classroom visits to observe student teachers in the field. They evaluated candidates' classroom teaching as part of formative assessment, but they did not assign grades for the field experience component. The program coordinator (elementary or secondary), rather than the supervisors, assigned grades for the student teaching component based on supervisor assessments, mentor teacher evaluations, lesson plans, professional conduct, and other assignments.

Data for this study were drawn from candidates' records in the teacher education program and included: (a) demographic and placement information; (b) scores and written comments on the PACT teaching event; (c) predicted scores for the PACT teaching event; (d) classroom observations by the supervisor; (e) mentor teacher evaluations; and (f) student transcripts. A conversation with the Director of Teacher Education served to clarify procedures and gather information about possible extenuating circumstances of candidates. The records provided to researchers included assigned case numbers to protect individual identities. In this article, I use pseudonyms for candidates rather than case numbers.

The study employed a case study design, which is particularly well suited to examining "how" and "why" contemporary events occur (Yin, 1994). The aim of this follow-up study was to examine how and why differences between predictions and total scores on the PACT teaching event occurred. The study identified and examined critical cases which had "strategic importance in relation to the general problem" (Flyvbjerg, 2001, p. 78). Given the study's focus on predictions of performance for low and high performers, the critical cases became those candidates with the largest differences between predictions and total scores. Data analysis

Predictions and Performance on the PACT Teaching Event

proceeded in three phases. The first phase focused on analysis of the individual cases. For each case, I identified the differences between predictions and scores for each of the 11 sections and examined the scorer's written comments to determine the justification for the scores. I similarly examined the classroom observation documents to determine the supervisor's assessment of the candidate's performance as a classroom teacher and to identify possible rationales for the predicted scores. For the five cases in which mentor teacher evaluations were available, I looked for corroborating or disconfirming evidence about the candidate's classroom teaching. I also reviewed student transcripts to identify candidates' grades in the methods courses and in student teaching. I used grades in methods courses because the curriculum and assignments for those courses are the most closely connected to classroom teaching. Candidates who are preparing to teach in elementary schools complete multiple methods courses including mathematics, science, language arts, social studies, reading, visual and performing arts, and physical education. For candidates preparing to teach at the secondary level, I reviewed grades for the subject-specific methods courses and for a course about reading and writing in secondary schools. Since the various courses are taught by different instructors, the grades provide evidence of how multiple individuals judged the candidate's performance. A candidate's grade for the student teaching component is based on a range of evidence including supervisor assessments, mentor teacher evaluations, lesson plans, professional conduct, and other assignments. For each individual case, I analyzed the combined sources of evidence to develop a potential explanation for the differing perspectives of the scorers and the supervisors.

In the second phase, I conducted cross-case comparisons within each group: the high performers and the low performers. I looked for group patterns within each data source and across data sources. For example, I compared candidates' undergraduate grade point averages to determine if low performance on the PACT teaching event correlated with low undergraduate performance. Similarly, I looked for patterns that would reveal if high or low performance could be attributed to tendencies of a particular scorer or supervisor. During this phase, I talked with the Director of Teacher Education to determine if candidates had extenuating circumstances that may have affected their performance in courses, student teaching, or the performance assessment. The third phase focused on comparisons across the two groups. I investigated patterns across the groups and examined evidence for emergent explanations for the discrepancies between predictions and scores. As a form of member checking, I presented the findings to a focus group that included university instructors and supervisors, graduate students, and teachers.

Findings

The findings section provides an overview of information about each group followed by narrative descriptions for each high and low performer. Each case includes

a bar chart that indicates the candidate's predicted and actual scores for the eleven questions, which are grouped in the five categories: planning, instruction, assessment, reflection, and academic language. Following the case narratives, the discussion section examines potential reasons for the differences between predictions and scores.

High Performers

Table 2 summarizes background information and predicted and actual PACT total scores for the six high performers. The narratives focus on the four high performers whose records included supervisor observations.

Grace. After completing both a bachelor's and a master's degree in civil engineering, Grace entered the single-subject credential program in mathematics. Her undergraduate grade point average was 3.0. She received a near-perfect total score on the PACT teaching event, 43 out of a possible 44. Her supervisor predicted a total score of 22, an under-prediction of nearly half the possible total. Whereas her supervisor predicted an acceptable level of performance (2 on each of the 11 sections), Grace received scores considered outstanding and rare for beginning teachers (4s on ten sections and a 3 on one section: Designing Assessments).

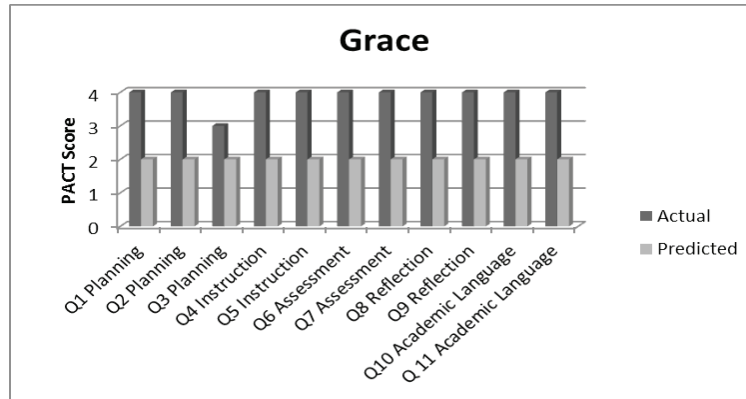
The scorer of Grace's teaching event commented repeatedly on the specificity and details included in Graces' plans, instructional strategies, and reflections. For example, the scorer wrote: "Proposed changes are specific: more hands-on activities, writing prompts to discover and monitor understanding, and introduction of controlled group work." The scorer also noted that Grace's "strategies for intellectual engagement [are] explicitly identified in commentary and the attention given is clearly reflected in

Table 2
High Performers

High Performer	Gender	Credential Program	Bachelors Degree	BA GPA	Predicted Score	PACT Score	Difference
Grace	F	Single-subject/ Mathematics	Civil Engineering	3.00	22	43	21 pts.
Caroline	F	Multiple-subject	Child Development	3.82	22	41	19 pts.
James*	M	Single-subject/ Science	Biological Sciences	2.88	22	39	17 pts.
Matthew*	M	Single-subject/ Mathematics	Chemical Engineering	3.09	21	38	17 pts.
Kathryn	F	Multiple-subject	Liberal Studies	3.11	25	39	14 pts.
Elizabeth	F	Single-subject/ Science	Biological Sciences	3.54	26	39	13 pts.

*Missing supervisor observations

Predictions and Performance on the PACT Teaching Event



video.” In monitoring student progress, Grace “focuses on specific learning needs of individuals and class” and her “adjustments are specifically targeted at how to further increase and deepen student learning to help meet the objectives.”

Her university supervisor’s comments from classroom observations indicated that Grace made steady progress and improvements from one visit to the next. The supervisor described Grace as “very organized and prepared” and pointed out how she adapted her lessons based on what had and had not worked previously. The comments highlighted how Grace’s questioning strategies improved, how teacher-student interaction increased, and how she gained rapport with the students. For example, in the second observation, the supervisor wrote: “The level of interaction between you and the students has increased and improved considerably since the first observation.” The supervisor concluded the first observation by expressing confidence in Grace’s ability to develop into an effective teacher: “I know that you will continue to improve and to develop professionally as a teacher. Already you have the traits of a good teacher.” In a subsequent observation, the supervisor noted, “Your questioning strategies have improved considerably, and students are responding well to them.” The supervisor also commented on how Grace spent time addressing students’ misconceptions rather than “rushing on ahead to finish your lesson in the designated time.” Following observations, the supervisor made a habit of noting improvements as well as making suggestions. Grace received a grade of A for each quarter of student teaching, an A- in her mathematics methods course, and an A in the reading and writing in secondary classrooms course. She successfully completed the credential program.

Caroline. After completing a bachelor’s degree in child development with a 3.8 grade point average, Caroline entered the multiple-subject credential program. Whereas her supervisor predicted an acceptable level of performance, Caroline received scores considered exceptional and rare for beginning teachers. Her supervisor predicted 2s on all sections for a total of 22. Caroline’s total score was 41

Judith Haymore Sandholtz

(4s on seven sections and 3s on Designing Assessments; Understanding Language Demands; and Supporting Academic Language Development).

The scorer’s comments highlighted evidence from Caroline’s videotaped segment and the comprehensive nature of her written commentary and submitted materials. For example, the scorer wrote:

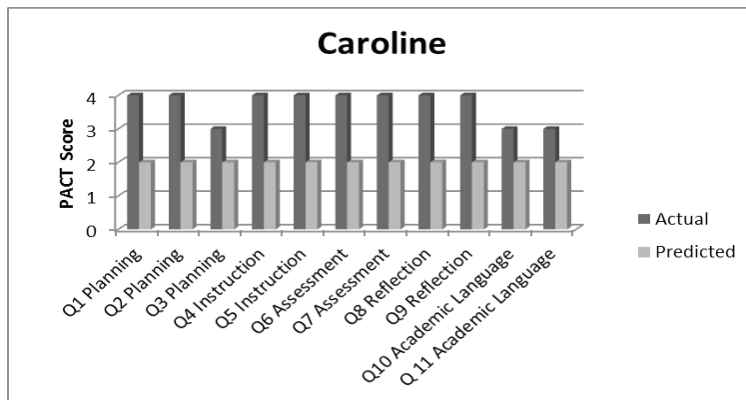
The candidate’s analysis of students’ work was clear and detailed. . .

During the video clip the candidate does a masterful job of questioning to elicit explanations of students’ reasoning. Input from students’ answers gave cause for several discussions. . .

In the section on establishing a balanced instructional focus, the scorer wrote:

Both learning tasks and the set of assessment tasks focus on multiple dimensions of mathematics learning through clear connections among computations/procedures, concepts, and reasoning/problem solving strategies. A progression of learning tasks and assessments guides students to build deep understandings of the central focus of the learning segment. The daily lesson plans include objectives and assessments that build a progressive understanding of the content. Multiple experiences are provided aimed at getting students to understand the meaning and various combinations.

Her university supervisor’s comments from classroom observations indicated that Caroline “is eager for feedback and accepts constructive criticism” and that she “works through dilemmas to improve her teaching.” The supervisor pointed out both effective strategies that Caroline implemented and areas for improvement. For example, the supervisor suggested more effective strategies for providing directions, handling a lack of time, and having students work with partners. The supervisor’s comments affirmed that Caroline took her responsibilities seriously, listened to feedback, and made progress in her teaching from one observation to the next. On candidate assessment forms, Caroline’s mentor teacher ranked her classroom experience very positively, indicating that she consistently used the



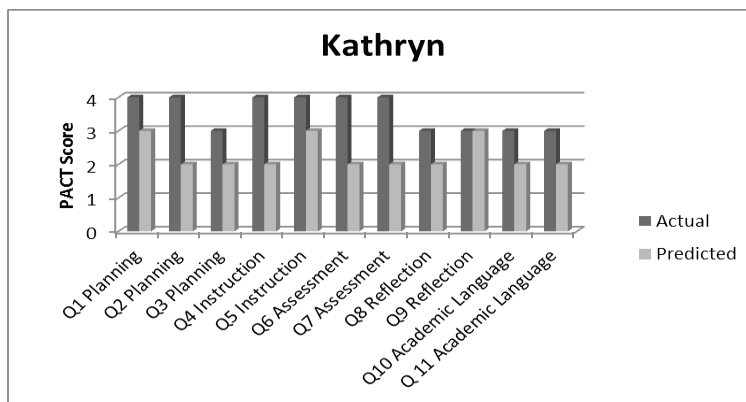
Predictions and Performance on the PACT Teaching Event

specified knowledge, skills, or practices in the various domains appropriately and competently. Caroline received a grade of A in all of the curriculum and methods course for elementary school, and grades of A and A+ for student teaching. She successfully completed the credential program and also earned a Master of Arts in Teaching (MAT degree).

Kathryn. Before entering the multiple-subject credential program, Kathryn completed a bachelor's degree in liberal studies with a 3.11 grade-point average. Her supervisor predicted that she would receive a passing score of 2 on eight of the eleven sections and an advanced score of 3 on three sections (Establishing a Balanced Instructional Focus, Monitoring Student Learning during Instruction, and Reflecting on Learning) for a total score of 25. Kathryn's scores were higher than predicted on all eleven sections. She received scores of 3 on five sections and scores of 4 on six sections, for a total score of 39.

In the planning sections, the scorer wrote that Kathryn's plans: a) demonstrate "a progression of learning tasks and assessments [that] guides students to build deep understandings of the central focus of the learning segment," b) draw on "students' prior learning as well as experiential backgrounds or interests to help students reach the learning segment's objectives," c) include "learning tasks [that] include scaffolding and other structured forms of support to provide access to objectives, and d) include "well-integrated instructional strategies that are tailored to address a variety of specific student learning." The scorer commented on instructional strategies evident in the teaching in the video clip that "are explicit, and clearly reflect attention to students with diverse characteristics, learning needs, and/or language needs." In addition, the candidate "monitors student understanding by eliciting student responses that require mathematical reasoning or problem solving strategies."

In each classroom observation, her supervisor documented "areas of best practices" and "areas to think about." In the first observation, the supervisor noted that Kathryn developed well-planned lessons, established a positive and inviting class-



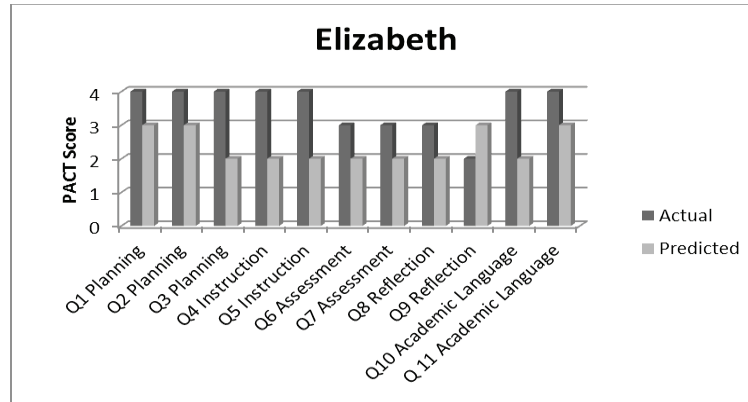
room environment, and knew the content well. The supervisor offered suggestions about managing student behavior and incorporating strategies that engage students at strategic points during the lesson and make students accountable. In subsequent observations, the supervisor noted improvement in classroom management and praised the way in which Kathryn integrated curricular areas and focused the students on the lesson. The supervisor commented on the pleasure in “watching [Kathryn] grow and apply the strategies we have discussed during this [student teaching] assignment.” Her mentor teacher ranked Kathryn’s classroom experience positively, indicating that she demonstrated developing ability or consistent use of specified knowledge, skills, or practices appropriately and competently. Her mentor teacher observed positive changes in classroom management and particularly appreciated how Kathryn listened to and acted upon feedback. She also noted that Kathryn was able to accurately reflect on her own strengths and areas for improvement. In the curriculum and methods courses for elementary school, Kathryn received an A+ in science, an A- reading, and an A in the others. She received an A+ for each quarter of student teaching. Kathryn successfully completed the credential program and also earned a Master of Arts in Teaching (MAT) degree.

Elizabeth. Sixteen years after completing a bachelor’s degree in biological science with a 3.54 grade point average, Elizabeth entered the single-subject credential program in science. In the interim, she completed a master’s degree in occupational therapy. Her supervisor predicted Elizabeth would receive a total score of 26 (2s on seven sections and 3s on four sections). But she achieved a much higher total score of 39 and earned an outstanding score of 4 on seven sections. On ten sections, she received a higher score than predicted; but on one section, Reflecting on Learning, she received a score of 2 whereas her supervisor predicted a score of 3.

The scorer indicated that the reflection commentary was missing, so the scorer referred instead to the lesson reflections. In comments, the scorer indicated that the candidate received scores of 4 because of the intentionality of the lesson progression in lesson plans, the number and quality of supports provided for students, the assessment modifications, the way in which the candidate elicited explanations of student thinking during instruction, and the explicit descriptions in the written commentary.

The supervisor’s comments from classroom observations indicated that Elizabeth enjoyed a natural rapport with students and used instructional strategies that involved teacher-student interaction. The supervisor noted that her instruction met academic content standards and included various strategies for making content accessible to students. The supervisor expressed some concern about the significant amounts of time Elizabeth spent designing lessons and creating most of her own materials and talked with the mentor teacher about “providing a more scaffolded experience” that would allow Elizabeth to focus on refining rather than developing lessons. The supervisor pointed out that if she spent less time on basic lesson design,

Predictions and Performance on the PACT Teaching Event



Elizabeth could work on making students accountable, monitoring students during instruction, and differentiating instruction to meet all students' needs. In addition to pointing out Elizabeth's strengths, the supervisor recommended strategies for improving student behavior during some class activities and offered subject-specific suggestions about organizing and facilitating science labs. After each observation, the supervisor concluded that Elizabeth was making good progress. Elizabeth received a grade of A- for each quarter of student teaching, A in her science methods course, and A in the reading and writing in secondary classrooms course. Elizabeth successfully completed the credential program.

Low Performers

Table 3 summarizes background information and predicted and actual PACT total scores for the four low performers.

Vera. Vera entered the multiple-subjects credential program with a bachelor's degree in psychology and a 3.0 undergraduate grade point average. Her supervisor predicted that Vera would do very well on the teaching event with a total score of 32. The supervisor predicted primarily 2s or 3s, and even predicted 4s on the three

Table 3
Low Performers

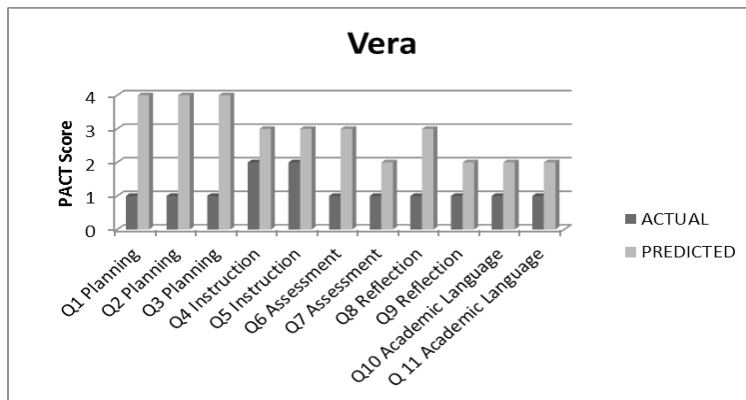
High Performer	Gender	Credential Program	Bachelors Degree	BA GPA	Predicted Score	PACT Score	Difference
Vera	F	Multiple-subject	Psychology	3.00	32	13	19 pts.
Martha	F	Multiple-subject	Sociology	2.99	31	16	15 pts.
Steven	M	Single-subject/ Mathematics	Computer Science	2.90	26	12	14 pts.
Susana	F	Multiple-subject	Liberal Studies	3.58	33	20	13 pts.

Judith Haymore Sandholtz

planning sections (Establishing a Balanced Focus, Making Content Accessible, and Designing Assessments). In contrast to the predictions, Vera received a total score of only 13. She received passing scores of 2 on only two sections (Engaging Students in Learning and Monitoring Student Learning during Instruction) and failing scores of 1 on the other nine sections. Whereas her supervisor predicted exceptional performance on the planning sections, her scores were not passing.

The scorer’s comments indicated that Vera submitted a very limited commentary, thereby inhibiting the scorer’s ability to judge the candidate’s understanding. For example, in the planning category, Vera submitted plans for only two lessons rather than three. The scorer noted that one lesson was well developed while the other was “extremely limited in content.” In another section, an analysis of a student work sample consisted of a single sentence, and an assessment section had been skipped entirely.

In classroom observations, the supervisor commented on Vera’s effective implementation of the behavior management system, the way in which she maintained clear expectations for students, and her ability to keep students both motivated and focused. In the classroom, Vera developed a professional teaching manner that was “very natural, commanding, and nurturing.” The supervisor pointed out numerous positive actions by Vera, but also offered suggestions related to specific instructional programs being used in math and language arts. For example, the supervisor noted math activities that help students develop deep understanding and encouraged Vera “to introduce and reinforce different ‘transitional’ vocabulary. On candidate assessment forms, her mentor teacher indicated that Vera was demonstrating a developing ability to use knowledge and skills identified in the performance standards appropriately and competently. She commented that Vera was always well prepared, had good classroom control, and “encouraged questions and comments from all learners.” By the final evaluation, the mentor teacher felt Vera had progressed to a consistent use of skills in the performance standards and noted that one of Vera’s greatest strengths was her ability “to self-reflect and modify



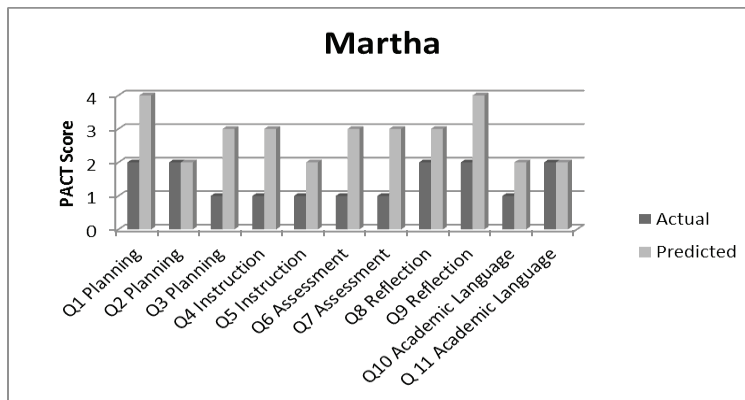
Predictions and Performance on the PACT Teaching Event

lessons to better meet the needs of all learners.” In the curriculum and methods courses for elementary school, Vera received an A in science, an A- in both reading and physical education, and a B in the mathematics course. She received grades of A and A- for student teaching. Vera passed the PACT teaching event on her next submission and successfully completed the credential program.

Martha. After completing a bachelor’s degree in sociology with a 2.99 undergraduate grade point average, Martha entered the multiple-subjects credential program. Martha’s total score of 16 (2s on five sections and 1s on six sections) was far below her supervisor’s prediction of 31. On two of the 11 sections (Making Content Accessible and Supporting Academic Language Development), the supervisor’s predictive score of 2 matched the actual score. But on six sections, there was a 2-pt range of difference, meaning the prediction was “advanced level of performance” but the score was “not passing” or the prediction was “outstanding” and the score was “adequate.”

The scorer’s comments pointed out a lack of connection between assessment tools and learning activities, a lack of explanation about how proposed strategies would improve students’ mathematical understanding, and limited opportunities for students to develop their own understandings. The scorer indicated that, when monitoring student learning during instruction, the candidate responded to students’ answers as correct or incorrect and did not ask them to justify or explain their thinking. In the section on analyzing student work from an assessment, the scorer noted that Martha again focused on what students did correctly or incorrectly, without discussing underlying misconceptions or different levels of student learning.

In classroom observations, the supervisor stated that Martha’s instruction met academic standards, included clear learning objectives, and was based on detailed, organized plans. Martha incorporated a range of instructional strategies and included innovative activities “to make learning compelling.” She developed a comfortable rapport with the students and had good classroom management. The supervisor



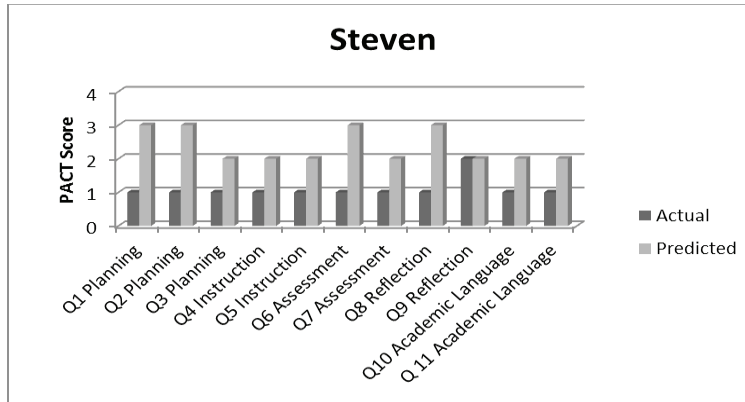
commented on Martha's progress each visit, her willingness to take advice and suggestions, and her "excellent job of reflecting and striving to improve her teaching." The supervisor suggested that "by continuing to make improvements to an already wonderful teaching base," Martha was "on her way to becoming a great teacher!"

Her mentor teacher ranked Martha's classroom experience positively, indicating that she demonstrated developing ability or consistent use of specified knowledge, skills, or practices appropriately and competently. In comments, the mentor teacher indicated that Martha had used a variety of instructional strategies and groupings, including both independent and collaborative learning. The mentor teacher pointed out that Martha had improved in lesson sequencing and classroom management techniques, and recommended additional practice in making adjustments during lessons to keep students engaged and differentiating instruction to meet student needs. In the curriculum and methods courses for elementary school, Martha received a B in the reading course and an A or A- in the others. She received grades of A- and A+ for student teaching. She passed the PACT teaching event on re-submission and successfully completed the program.

Steven. Before becoming a candidate in the single-subject mathematics credential program, Steven completed a bachelor's degree in computer science with a 2.9 grade point average. Steven's supervisor predicted that he would receive a total score of 26 on the PACT teaching event. The predicted scores for the individual sections were primarily 2s, but on four sections (Establishing a Balanced Instructional Focus, Making Content Accessible, Analyzing Student Work from an Assessment, and Monitoring Student Progress), the supervisor predicted a score of 3, an advanced level of performance. Steven's performance was far below the supervisor's predictions. The prediction and score matched on only one section: a 2 on Reflecting on Learning. Steven received failing scores of 1 on the other ten sections for a total score of only 12. The scorer noted that Steven selected two disjointed lessons for his teaching event rather than a learning segment with a clear focus, and throughout his commentary, he offered superficial analysis with insufficient details. For example, the scorer indicated that the candidate, in analyzing student work, "just merely stated what the students could or could not do and again lacked detail with each student's analysis consisting of just two or three sentences. An analysis at such a superficial level will not help inform the candidate on how to adapt in future lessons."

Corresponding with the predicted scores in planning, the university supervisor noted in observation reports that Steven had good lesson plans. After one of the early observations, the supervisor wrote that Steven had "a wonderful plan to allow students to think deeply about an interesting math problem" and that his "planning gives me good confidence and hope that he will become a great math teacher." However, Steven's implementation of his instructional plans was not as successful. As both he and his supervisor reflected, he struggled to maintain student

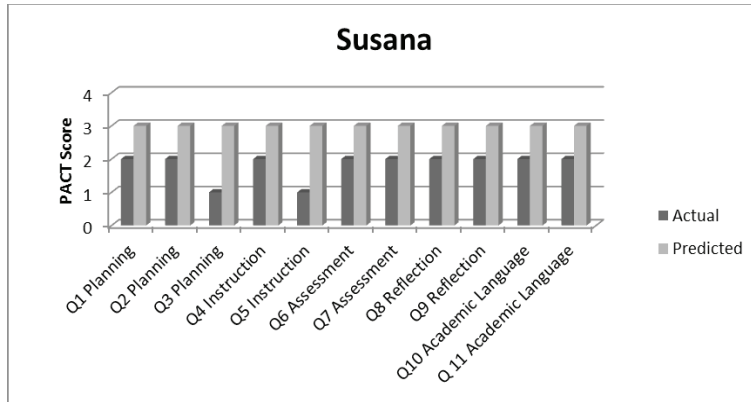
Predictions and Performance on the PACT Teaching Event



attention and did not make effective use of time. The supervisor suggested changes he could make and concluded that Steven “just needs time to learn how to be successful in implementing the lessons.” In a subsequent observation, the supervisor again noted that “the lesson was very well-planned” with instructional strategies that link to students’ prior knowledge and that should “catch their interest.” However, classroom management issues became “a major struggle.” The supervisor provided specific suggestions in response to the observed problems, emphasizing the need to communicate expectations clearly and to consistently and immediately enforce consequences. The supervisor noted that Steven had a “tough group of students” but would “learn much about managing a class from this challenge.” As the year progressed, Steven continued to plan lessons that “had novelty, allowed for autonomy, and met different levels of learning” and the supervisor pointed out significant improvement in classroom management.

Steven received a grade of B in the mathematics methods course, an A in the reading and writing for secondary schools course, and a grade of B for each quarter of student teaching. According to the Director of Teacher Education, by the time Steven finished student teaching, he questioned whether he still wanted to be a teacher. However, he passed the PACT teaching event on the next submission and went on to complete a post-credential Master of Arts in Teaching degree.

Susana. Two years after completing a bachelor’s degree in liberal studies with a grade point average of 3.58, Susana began the multiple-subjects teacher credential program. Her supervisor predicted that Susana would do well on the performance assessment and receive a total score of 33. The supervisor predicted a score of 3, an advanced level of performance, on each of the 11 sections. Susana’s total score was 20. She received a passing score of 2 on nine sections and a failing score of 1 on two sections: Designing Assessments and Monitoring Student Learning during Instruction. For those sections, the scorer noted that some of the assessments were not matched to the lesson content and that the candidate primarily monitored student understanding



by asking surface-level questions. In other sections, the scorer indicated that there were “vague connections among the concepts and procedures,” that the analysis “focused on what students could do or couldn’t do” and that “more differentiation is needed.”

Susana’s supervisor and mentor teacher found much to praise in their evaluations of Susana’s teaching. In the initial observation, the supervisor commented on Susana’s good control of the classroom, her strong rapport with students, and her use of strategies that allow her to check for understanding. The supervisor also made specific recommendations related primarily to scaffolding. In subsequent evaluations, the supervisor mentioned her well-organized lessons, her clear expectations for students, and her use of positive reinforcement. The supervisor described Susana’s lessons as “impressive” and commented on Susana’s overall good job in the classroom. Susana’s mentor teacher noted that Susana was consistently well-prepared, had strong classroom management skills, and created a positive learning environment with high student engagement. She pointed out that Susana studied the subject matter in advance in order to make it comprehensible to students and considered “students’ level in order to differentiate learning.” Susana’s organizational skills allowed her to “utilize every minute for learning.” Susana had high performance in her methods courses and student teaching. In the curriculum and methods courses for elementary school language arts, mathematics, and science, Susana received grades of A+. She earned grades of A in the other methods courses and a grade of A+ for each quarter of student teaching. Susana successfully passed the PACT teaching event on re-submission and completed her teaching credential and a Master of Arts in Teaching (MAT) degree.

Discussion

In looking across the cases to understand why these differences between predictions and scores occurred, I found no distinct patterns related to particular supervisors or scorers, undergraduate performance, or grades in methods courses.

Predictions and Performance on the PACT Teaching Event

In this program, the supervisors complete the PACT training each year and must pass the established PACT calibration standard to serve as scorers. Given their familiarity with the format, requirements, and standards of the PACT teaching event as scorers, the differences between predictions and scores would not arise from lack of knowledge about the performance assessment itself. The high performers, Grace, Caroline, Kathryn, and Elizabeth, all had different scorers for their PACT teaching events. In addition, they all had different university supervisors. Similarly, the low performers, Vera, Martha, Steven, and Susana, all had different scorers and university supervisors. Consequently, the differences also do not appear to be due to tendencies of a particular scorer or supervisor.

Cases from the larger group of high and low performers, in which the supervisor/scorer pairs were the same, provide additional evidence that the differences do not result from some individuals being “easier graders” than others. For example, one high performer had both a prediction and a total score of 39; but another high performer with the same supervisor/scorer pair had a prediction of 21 and a total score of 38. This example illustrates that the same supervisor anticipated that one candidate would excel on the PACT and the other candidate would barely pass, yet both received high total scores on the PACT. Similarly, in the cases of two low performers with the same supervisor/scorer pair, the prediction and score matched in one case but differed by 10 points in the other case. If a supervisor were consistently predicting higher scores, the range between predictions and scores for the same supervisor/scorer pairs should be similar across candidates. When I examine the cases with the greatest differences between predictions and scores and the cases of high and low performers with the same supervisor/scorer pairs, I don’t find patterns suggesting that differences reflect grading tendencies of particular supervisors or scorers.

In addition, there is not a clear pattern related to undergraduate performance. Both groups show a similar range in undergraduate grade point averages: about 2.9 to 3.8 for the high performers and 2.9 to 3.6 for the low performers. Similarly, there is not a definitive pattern related to candidates’ grades in methods courses or student teaching. The candidates in the high-performing group had high grades in their courses. However, the candidate with the highest grades, Susana, and the candidate with the lowest grades, Steven, were both low performers. Susana consistently received very high grades across courses, which indicates that a number of different instructors viewed her work as outstanding. She received grades of A+ in three methods courses and grades of A in the others. In addition, she received grades of A+ for each quarter of student teaching. Steven received a grade of B in his mathematics methods course and for each quarter of student teaching. In this program, a grade of B- or below is considered unsatisfactory. As described earlier, university supervisors do not assign grades for student teaching in this program. The program coordinator assigns the student teaching grades by drawing upon a range of evidence such as supervisor assessments, mentor teacher evaluations, lesson plans, seminar assignments, and professional conduct.

The discrepancies between the predictions and scores appear to stem from three differences in the tasks of supervisors and scorers. First, supervisors and scorers draw upon different data sources. Whereas supervisors make predictions based on classroom observations and formative assessments, scorers make judgments based on teaching artifacts and written commentaries. Supervisors in this program are not directly involved in preparing candidates for the performance assessment and do not review drafts of written commentaries. Their predictions stem from their observations of classroom teaching and discussions with candidates about their plans and instructional practice, but not from candidates' written analyses of their teaching. Candidates who may be effective classroom teachers may not be as skilled in writing about their instructional practice. In addition, some candidates may devote more attention to either their classroom teaching or the performance assessment. For example, Grace and Caroline, two high performers, submitted written commentaries that were comprehensive, specific, and detailed. In contrast, Vera submitted a written commentary that was extremely limited and missing some components. The Director of Teacher Education remembered Vera as a highly capable student who was assuming significant personal responsibilities due to her mother's health. At the time of the performance assessment, she may have faced competing responsibilities and assigned greater priority to student teaching.

Second, supervisors observe and gauge progress over time, while scorers make a single judgment at one point in time. The university supervisors in this program focus on formative evaluation and feedback; they do not assign grades nor make summative judgments. In contrast, scorers focus on making a summative assessment. As evidenced in the case studies, the comments of the university supervisors demonstrated a progression in each candidate's teaching over time. The supervisors tended to anticipate professional growth in the candidates over the course of their student teaching experience. For the high performers, the supervisors noted effective strategies and offered constructive criticism and suggestions. They predicted that the candidates would definitely pass the performance assessment, but they didn't predict exceptional performance. The supervisors' predictions may reflect where they viewed the candidates in an overall trajectory of professional growth.

Third, supervisors assess candidates' teaching in active classrooms with changing situations whereas scorers view a bounded, pre-selected segment of a class. What supervisors view in observations may not correspond with what scorers view in the PACT teaching event. For example, Steven's supervisor repeatedly noted significant problems with classroom management, but the scorer referred to classroom management only once. The scorer wrote that "classroom management was slightly problematic" but did not interfere with learning. The most significant problem in Steven's PACT teaching event was the limited and superficial analysis. Vera's supervisor observed a well-planned, competent student teacher in the classroom, but the scorer encountered an incomplete, substandard teaching event. In Grace's case, the supervisor saw a student teacher who was making steady progress and

Predictions and Performance on the PACT Teaching Event

improvement over time whereas the scorer viewed a well-developed lesson segment with an exceptional level of analysis. As these case studies highlight, candidates who struggle or demonstrate competence in the classroom may excel or flounder on the teaching performance assessment.

Conclusion and Implications

The findings of this study support the recommendation that multiple methods are needed to provide a comprehensive assessment of candidates' progress. Candidates who fail, or excel, on a performance assessment, such as the PACT teaching event, may appear more, or less, competent in overall assessments using multiple data sources. In this study, three candidates who failed the performance assessments demonstrated competence in courses and student teaching, as evidenced by grades, supervisor observations and mentor teacher evaluations. The other candidate who failed the assessment exhibited strong planning skills but struggled with other aspects of student teaching. The candidates with particularly high scores on the performance assessment clearly demonstrated overall competence in student teaching but were not identified as extraordinary. As these cases of high and low performers highlight, performance assessments and supervisor perspectives may offer varying views of a candidate's skills and progress. When multiple sources of evidence are combined, a more comprehensive assessment of a candidate's competence emerges.

The role of performance assessments in credentialing decisions is a key issue for teacher education programs for reasons that extend beyond resource requirements. First, if performance assessments become a single, high-stakes measure of teacher education outcomes and pre-service teacher qualifications, multiple perspectives will be lost. Different strategies that are used to measure effective teaching highlight different aspects of teacher quality (Peterson, 1987, 2000). Without multiple sources of information about candidates' effectiveness, our overall judgments about pre-service teachers' abilities may favor some aspects of teacher quality over others. Second, researchers question if the use of performance assessments for high-stakes credentialing decisions will mediate the influence of the assessment on teacher learning; for example, candidates may be less open about their weaknesses if their credential could be affected (Chung, 2008). Third, if credentialing decisions rely solely on a performance assessment, pre-service candidates may shift their focus from making steady progress in student teaching to producing an exemplary teaching event portfolio. The components of the performance assessment could end up taking priority over day-to-day teaching in the classroom.

Performance assessments such as PACT offer a valuable form of evidence about candidates' performance on authentic teaching tasks and focus pre-service candidates' attention on student learning. In addition, the assessments have the potential to serve as professional learning experiences and promote teacher reflection on practice (Bunch, Aguirre, & Tellez, 2009; Darling-Hammond & Snyder,

2000; Okhremtchouk et al., 2009). Performance assessments also can inform teacher education programs about areas of strength and areas for improvement (Darling-Hammond, 2006; Pecheone & Chung, 2006). However, the value of performance assessments may be undermined if they become a single, high-stakes measure of candidates' competence and teacher education outcomes. Given the complexity of teaching and learning, we need strategies that "provide a variety of lenses on the process of learning to teach" (Darling-Hammond, 2006, p.120). To make comprehensive assessments of pre-service teachers' abilities and progress, teacher education programs stand to benefit from drawing upon multiple methods and sources of data. Moreover, the potential benefits of performance assessments may be best realized when they are used in combination with other strategies.

References

- Arends, R. I. (2006a). Performance assessment in perspective: History, opportunities, and challenges. In S. Castle and B.S. Shaklee (Eds.), *Assessing teacher performance: Performance-based assessment in teacher education* (pp. 3-22). Lanham, MD: Rowman & Littlefield Education.
- Arends, R. I. (2006b). Summative performance assessments. In S. Castle & B. S. Shaklee (Eds.), *Assessing teacher performance: Performance-based assessment in teacher education* (pp. 93-123). Lanham, MD: Rowman & Littlefield Education.
- Bunch, G. C., Aguirre, J. M., & Tellez, K. (2009). Beyond the scores: Using candidate responses on high stakes performance assessment to inform teacher preparation for English learners. *Issues in Teacher Education*, 18(1), 103-128.
- California Commission on Teacher Credentialing. (2006). *Summary of commission responsibilities for major provisions of SB 1209*. Retrieved from <http://www.ctc.ca.gov/educator-prep/SB1209/default.html>
- Chung, R. R. (2008). Beyond assessment: Performance assessments in teacher education. *Teacher Education Quarterly*, 35(1), 7-28.
- Darling-Hammond, L. (1986). A proposal for evaluation in the teaching profession. *Elementary School Journal*, 86(4), 531-551.
- Darling-Hammond, L. (2006). Assessing teacher education: The usefulness of multiple measures for assessing program outcomes. *Journal of Teacher Education*, 57(2), 120-138.
- Darling-Hammond, L., & Snyder, J. (2000). Authentic assessment of teaching in context. *Teaching and Teacher Education*, 16(5-6), 523-545.
- Darling-Hammond, L., & Sykes, G. (1999). *Teaching as the learning profession: Handbook of policy and practice*. San Francisco: Jossey-Bass.
- Delandshere, G., & Arens, S. A. (2001). Representations of teaching and standards-based reform: Are we closing the debate about teacher education? *Teaching and Teacher Education*, 17, 547-566.
- Flyvbjerg, B. (2001). *Making social science matter*. Cambridge, UK: Cambridge University Press.
- Guaglianone, C. L., Payne, M., Kinsey, G. W., & Chiero, R. (2009). Teaching performance assessment: A comparative study of implementation and impact amongst California State University Campuses. *Issues in Teacher Education*, 18(1), 129-148.
- Haertel, E. H. (1991). New forms of teacher assessment. In G. Grant (Ed.), *Review of*

Predictions and Performance on the PACT Teaching Event

- research in education* (Vol 17, pp. 3-29). Washington, DC: American Educational Research Association.
- Leinhardt, G. (2001). Instructional explanations: A commonplace for teaching and location for contrast. In V. Richardson (Ed.), *Fourth handbook of research on teaching* (pp. 333-357). Washington, DC: American Educational Research Association.
- 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.
- National Board for Professional Teaching Standards. (1999). *What teachers should know and be able to do*. Arlington, VA: Author.
- Okhremtchouk, I., Seiki, S., Gilliland, B., Atch, C., Wallace, M., & Kato, A. (2009). Voices of pre-service teachers: Perspectives on the Performance for California Teachers (PACT). *Issues in Teacher Education, 18*(1), 39-62.
- PACT Consortium. (2009). *Implementation handbook*. Retrieved from http://www.pactpa.org/_main/hub.php?page=Implementation_Handbook
- Pecheone, R. L., & Chung, R. R. (2006). Evidence in teacher education: The Performance Assessment for California Teachers (PACT). *Journal of Teacher Education, 57*(1), 22-36.
- Pecheone, R. L., Pigg, M. J., Chung, R. R., & Souviney, R. J. (2005). Performance assessment and electronic portfolios: Their effect on teacher learning and education. *The Clearing House, 78*(4), 164-176.
- Peterson, K. (1987). Teacher evaluation with multiple and variable lines of evidence. *American Educational Research Journal, 24*(2), 311-317.
- Peterson, K. (2000). *Teacher evaluation: A comprehensive guide to new directions and practices* (2nd ed.). Thousand Oaks, CA: Corwin Press.
- Porter, A., Youngs, P., & Odden, A. (2001). Advances in teacher assessments and their use. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed.) (pp. 259-297). Washington, DC: American Educational Research Association.
- Richardson, V., & Placier, P. (2001). Teacher change. In V. Richardson (Ed.), *Fourth handbook of research on teaching* (pp. 905-947). Washington, DC: American Educational Research Association.
- Sandholtz, J. H., & Shea, L. M. (2012). Predicting performance: A comparison of university supervisors' predictions and teachers candidates' scores on a teaching performance assessment. *Journal of Teacher Education, 63*(1), 39-50.
- Shulman, L. S. (1986). Paradigms and research programs in the study of teaching: A contemporary perspective. In M. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed.) (pp.3-36). New York: Macmillan.
- Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57*, 1-22.
- Snyder, J. (2009). Taking stock of performance assessments in teaching. *Issues in Teacher Education, 18*(1), 7-11.
- Tellez, K. (1996). Authentic assessment. In J. Sikula (Ed.), *The handbook of research in teacher education* (2nd Ed.) New York: Macmillan.
- Yin, R. K. (1994). *Case study research: Design and methods* (2nd ed.). Thousand Oaks, CA: Sage.
- Zeichner, K. M. (2003). The adequacies and inadequacies of three current strategies to recruit, prepare, and retain the best teachers for all students. *Teachers College Record, 105*(3), 490-519.