Stanford Center for Opportunity Policy in Education
Stanford Center for Assessment, Learning and Equity

Developing and Assessing Beginning Teacher Effectiveness: The Potential of Performance Assessments

Linda Darling-Hammond,
Stephen P. Newton,
& Ruth Chung Wei

Developing and Assessing Beginning Teacher Effectiveness: The Potential of Performance Assessments

Linda Darling-Hammond, Stephen P. Newton, & Ruth Chung Wei

Stanford University

Abstract: The Performance Assessment for California Teachers (PACT) is an authentic tool for evaluating prospective teachers by examining their abilities to plan, teach, assess, and reflect on instruction in actual classroom practice. The PACT seeks both to measure and develop teacher effectiveness, and this study of its predictive and consequential validity provides information on how well it achieves these goals. The research finds that teacher candidates' PACT scores are significant predictors of their later teaching effectiveness as measured by their students' achievement gains in both English language arts and mathematics. Several subscales of the PACT are also influential in predicting later effectiveness: These include planning, assessment, and academic language development in ELA, and assessment and reflection in mathematics. In addition, large majorities of PACT candidates report that they acquired additional knowledge and skills for teaching by virtue of completing the assessment. Candidates' feelings that they learned from the assessment were strongest when they also felt well-supported by their program in learning to teach and in completing the assessment process.

Introduction

As teaching quality has become a major focus of policy attention, there is growing interest in improving teacher evaluation methods so that they both distinguish more readily among teachers with varying levels of skill, and so that they are more clearly associated with teachers' abilities to promote student learning. These concerns are as important for teacher assessment at the beginning of the career as they are for personnel evaluation on-the-job. Indeed, changing on-the-job evaluation will not by itself transform the quality of teaching. For all of the attention currently focused on identifying and removing poor teachers, it will be difficult to improve the quality of the profession if there is not also an strong supply of entering teachers who are well-prepared and able to continue to learn from practice.

One potentially promising approach to evaluating the quality of beginning teachers is the development of new performance assessments for teacher licensing that can both assess readiness to teach and, some research suggests, leverage improvements in preparation as well. After the National Board for Professional Teaching Standards created a new performance-based approach for assessing veteran teachers in the early 1990s, several states -- including California, Connecticut, and Oregon -- created performance assessments for beginning teacher licensure. Building further on the work in California, a recently-formed group of 27 states has formed a Teacher Performance Assessment Consortium to develop a nationally available assessment that can be used for purposes of initial licensure and program accreditation across the country (Darling-Hammond, 2010).

All of these performance assessments are portfolios that collect evidence of teachers' actual instruction, through videotapes, curriculum plans, and samples of student work and learning, along with teacher commentaries explaining the basis for teachers' decisions about what and how they

taught, in light of their curriculum goals and student needs, and how they assessed learning and gave feedback to individual students.

As this work progresses, it is important to evaluate both the predictive and consequential validity of these new assessments. Are candidate scores on the assessments associated with other evidence of their later effectiveness in the classroom? Does the use of the assessments support teacher learning? Does it provide useful information to preparation and induction programs about how better to support and strengthen teachers' practice?

This article contributes to this needed body of research by reporting results of early predictive and consequential validity studies for the Performance Assessment for California Teachers (PACT). We use linked student - teacher data from three large school districts in California to examine teacher PACT scores in relation to student learning gains. In addition, we use survey data from candidates involved in the PACT pilots to examine their self-reported learning from the assessment process and the extent to which this learning is related to their preparation context.

Below, we describe the PACT assessment, review prior studies on the PACT and other teacher performance assessments, describe the data base and methodology for the study, and report our results. Finally, we discuss the implications of this and related studies for the field of teacher assessment, and for future research.

The Performance Assessment for California Teachers

The Performance Assessment for California Teachers (PACT) measures beginning teachers' abilities to plan, implement, and assess instruction in actual classrooms while candidates are completing student teaching or an alternative route internship (Pecheone & Chung, 2006). It was developed beginning in 2002 by a consortium of 12 universities (all of the University of California

campuses, two California State University campuses, and two private universities). The PACT consortium, which has now grown to 31 university and teacher preparation district programs,[1] has been implementing the PACT for a decade. The consortium, coordinated by Stanford University, a participating institution, has supported ongoing refinement of the instrument, reliability and validity studies, training for scorers, audits of scoring reliability, and conferences for participating programs to share their curriculum and instructional strategies and learn from each other how to better support their candidates. In late 2007, the PACT assessment system was reviewed and approved by the California Commission on Teacher Credentialing (CCTC) for use as a state licensing requirement.

Such a performance-based measure responds to the call by the National Research Council (Mitchell, Robinson, Plake & Knowles, 2001) to develop broader assessments of teacher candidates, including performance in the classroom, and to validate them in terms of teachers' success in teaching. In contrast to paper and pencil measures of teacher knowledge or thinking, performance assessments provide a much more direct evaluation of teaching ability (Pecheone & Chung, 2006). Such assessments also have the potential to provide formative information for the candidates themselves and for teacher education programs, as they have the opportunity to examine a rich variety of products reflecting each teacher candidate's performance.

PACT consists of two classes of assessments, embedded signature assessments to be completed throughout the preparation program (for example, child case studies, curriculum units and other major learning experiences in teacher education), and a summative assessment of teaching knowledge and skills during student teaching, known as the teaching event (TE) (Pecheone and

[1] PACT members currently include UC-Berkeley, UC-Davis, UC-Irvine, UCLA, UC-Riverside, UC-San Diego, UC-Santa Barbara, UC-Santa Cruz; Cal Poly – SLO, CSU-Channel Islands, CSU-Chico, CSU-Dominguez Hills, CSU-Monterey Bay, CSU-Northridge, Humboldt State, Sacramento State; San Diego State; San Francisco State, San Jose State, Sonoma State; Antioch University Santa Barbara, Holy Names University, Mills College, Notre Dame de Namur University, Pepperdine University, St. Mary's College of California, Stanford, University of the Pacific, University of San Diego, USC; and the San Diego City Schools Intern Program.

Chung, 2006). This study evaluates the scores on the teaching event component of the PACT. In practice, the TE involves the following activities:

> "To complete the TE, candidates must plan and teach a learning segment (i.e., an instructional unit or part of a unit), videotape and analyze their instruction, collect student work and analyze student learning, and reflect on their practice." (Pecheone & Chung, p. 24)

More specifically, candidates plan a curriculum unit that addresses state learning standards and includes appropriate differentiation for English learners and students with disabilities. They describe their teaching context and rationale for the content and methods they have chosen. They teach a 3 to 5 day segment of the unit, writing reflections each evening on what students learned and what adjustments are needed for the next day. They submit a 15-minute continuous video clip from that period of time, writing a commentary about what the clip illustrates about their plans, decisions, teaching practice, and student learning. And they submit a set of student work from the class, with in-depth analysis of student learning and a reflection on what additional teaching is needed to support achievement of the learning goals for particular individuals and groups of students.

The work is assembled in a portfolio and submitted for assessment. Candidate work is then rated by trained and calibrated raters (teacher educators and teachers in the same teaching field) on a set of subject-specific rubrics that evaluate: Planning, Instruction, Assessment, Reflection, and Academic Language. Within these areas, the analytic scoring scheme is further shaped by a set of guiding questions, as the following example shows for elementary English language arts:

Planning

> EL1: How do the plans structure student learning of skills and strategies to comprehend and/or compose text?

EL2: How do the plans make the curriculum accessible to the students in the class?

EL3: What opportunities do students have to demonstrate their understanding of the standards/objectives?

Instruction

EL4: How does the candidate actively engage students in their own understanding of skills and strategies to comprehend and/or compose text?

EL5: How does the candidate monitor student learning during instruction and respond to student questions, comments, and needs?

Assessment

EL6: How does the candidate demonstrate an understanding of student performance with respect to standards/objectives?

EL7: How does the candidate use the analysis of student learning to propose next steps in instruction?

Reflection

EL8: How does the candidate monitor student learning and make appropriate adjustments in instruction during the learning segment?

EL9: How does the candidate use research, theory, and reflections on teaching and learning to guide practice?

Academic Language

EL10: How does the candidate describe student language development in relation to the language demands of the learning tasks and assessments?

EL11: How do the candidate's planning, instruction, and assessment support academic language development?

Raters are trained and audited, producing high levels of consistency in scoring, as documented in reliability studies (Pecheone and Chung, 2006). A set of validity studies conducted of the assessment over several years has informed ongoing refinements in the assessment instrument and scoring process (Pecheone & Chung, 2007).

Rationale for the Study

The present study, linking pre-service or intern teachers' performance on the PACT with their early career effectiveness, as measured by value-added assessment of their students' achievement, addresses the important issue of the predictive validity of the assessments. It follows up on an earlier, smaller study, which tracked the value-added scores of students of a small (n=14) cohort of teachers in San Diego during their first two years in the classroom. The teachers were part of an internship program (California's alternative route) preparing elementary teachers for bilingual classrooms. In this early pilot study, the PACT literacy portfolio scores of these new teachers were found to predict their students' gains on state ELA tests (Newton, 2010).

This article reports on a somewhat larger-scale study, using data from candidates in multiple teacher education programs hired to teach in three cities in California. The study seeks to evaluate whether this positive relationship holds up in different contexts and to examine whether certain sub-scores of the PACT measuring different dimensions of teaching are more predictive of teacher effectiveness than others.

Examining the ability of performance assessments to predict teachers' future effectiveness is important for several major reasons. First, it is important to validate PACT performance as a measure of teacher quality by relating it to other measures of teacher quality and effectiveness, such as value-added measures of their classroom performance. This kind of predictive validity study,

rarely pursued for most teacher tests, can provide greater confidence that the assessment is measuring aspects of teaching that contribute to student learning.

Furthermore, the use of a validated teacher performance assessment for teacher licensure allows a more timely decision about readiness for entry than direct measurement of value-added scores could provide (even if these scores were able to be used, appropriately, for later evaluation).

Third, the link between PACT performance and teacher effectiveness also may provide critical information for teacher education institutions about their own effectiveness. Performance assessment of preservice teachers can provide important advantages over tracking performance of program graduates in the field, because of their relative ease, timing, and rich feedback information organized around specific dimensions of teaching that programs can address in their curricula and clinical experiences.

Finally, leading measurement experts have suggested that developers and policymakers should be concerned about the consequential validity of assessments; that is, what effect the assessments have on learning and improvement for both test-takers and faculty or organizations that receive the results. This goal is important for performance assessments like the PACT, which have a dual purpose, explicitly intending to help develop competence as well as measure it. To evaluate their success, it is important to collect evidence about whether candidates perceive that they have learned about teaching from their participation in the assessment and whether programs have improved as a result of their participation in the process and their examination of the data. We treat the first of these questions, regarding candidate learning, in this study.

Review of the Literature

For many decades, teachers' scores on traditional paper-and-pencil tests of basic skills and subject matter, while useful for establishing academic standards, have failed to register a significant

relationship to their students' learning gains in the classroom (Andrews, Blackmon & Mackey, 1980; Ayers & Qualls, 1979; Haney, Madaus, & Kreitzer, 1986; Wilson et al., 2007). By contrast, well-designed performance-based assessments have been found to measure aspects of teaching related to teachers' effectiveness, as measured by student achievement gains. In addition, some studies indicate that the process of completing the assessments can stimulate teacher learning and that feedback from the assessments can support both candidate and program learning.

Relationships between Performance-Based Assessment Scores and Student Learning Gains

The longest standing such assessment is the portfolio used for National Board Certification, which has given rise to a number of studies, most of which have found positive influences on student learning gains. For example, Cavaluzzo (2004) examined mathematics achievement gains for nearly 108,000 high school students over four years in the Miami-Dade County Public Schools, controlling for a wide range of student and teacher characteristics (including experience, certification, and assignment in field, as well as Board certification). Each of the teacher quality indicators made a statistically significant contribution to student outcomes. Students who had a typical NBC teacher made the greatest gains, exceeding gains of those with similar teachers who had failed NBC or had never been involved in the process. The effect size for National Board Certification ranged from 0.07 to 0.12, estimated with and without school fixed effects. Students with new teachers who lacked a regular state certification, and those who had teachers whose primary job assignment was not mathematics instruction made the smallest gains.

Goldhaber and Anthony (2005), using three years of linked teacher and student data from North Carolina representing more than 770,000 student records, found the value-added student achievement gains of NBCTs were significantly greater than those of unsuccessful NBCT candidates and non-applicant teachers. Students of NBCTs achieved growth exceeding that of

students of unsuccessful applicants by about 5% of a standard deviation in reading and 9% of a standard deviation in math.

In two other large-scale North Carolina-based studies using administrative data at the elementary and high school levels, Clotfelter, Ladd, and Vigdor (2006, 2007) found positive effects of National Board certification on student learning gains, along with positive effects of other teacher qualifications, such as a license in the field taught. Comparing NBC teachers to all others (rather than to those who had attempted and failed the assessment, where the differences are greatest in most studies), they found effect sizes of .02 to .05 across different content areas and grade levels, with fairly consistent estimations using student and school fixed effects.

Using randomized assignment of classrooms to teachers in Los Angeles Unified School District, Cantrell, Fullerton, Kane, and Staiger (2007) found that students of NBC teachers outperformed those of teachers who had unsuccessfully attempted the certification process by 0.2 standard deviations, about twice the differential that they found between NBC teachers and unsuccessful applicants from a broader LAUSD sample not part of the randomized experiment, but analyzed with statistical controls.

Significant positive influences of NBC teachers on achievement were also found in much smaller studies by Vandevoort, Amrein-Beardsley, and Berliner (2004) and Smith, Gordon, Colby, and Wang (2005). Smith and colleagues also examined how the practices of their 35 NBCTs compared to those of 29 who had attempted but failed certification, finding significant differences which reflected the ways in which NBCTs fostered deeper understanding in their instructional design and classroom assignments.

Not all findings have been as clearly positive. Using an administrative data set in Florida, Harris and Sass (2007) found that NBC teachers appeared more effective than other teachers in

some but not all grades and subjects--and on one of the two different sets of tests evaluated (the Florida Comprehensive Assessment Test and the SAT-9). This study did not compare NBCs to those who had attempted certification unsuccessfully, which is the strongest comparison for answering the question of whether the Board's process differentiates between more and less effective teachers. Finally, using a methodology different than that used in most other studies, Sanders, Ashton, and Wright (2005) found effect sizes for National Board certified teachers similar to those of other studies (about .05 to .07 in math), but most of the estimates were not statistically significant because of the small sample sizes.

Other teacher performance assessments have been examined as well. For example, beginning teachers' ratings on the Connecticut BEST assessments - taken in the second year of teaching as the basis for a professional license -- were found to predict their students' achievement gains on state reading tests. The measure used was the Degrees of Reading Power test. The study used hierarchical modeling to isolate the effects of teachers nested within schools. Meanwhile other measures of teacher quality -- such as the selectivity of undergraduate college attended and scores on the Praxis subject matter tests -- had no influence on student gains (Wilson et al., 2007). In this study, a one-unit change in the portfolio score (on a 4-unit scale) was associated with a difference of about 4 months of learning time in an average year for the students in this study.

Similarly, as noted earlier, a small pilot study of the California PACT found that literacy portfolio scores of new intern teachers strongly predicted their students' gains on state ELA tests using four different value-added models (Newton, 2010). Sub-scores for the *assessment* dimension of the PACT (evaluating candidates' ability to use assessment data to support student learning) were particularly strongly related to student gains.

The relationship between PACT scores and student learning gains was substantial in this study: For each additional point a teacher scored on PACT (evaluated on a 44-point scale, based on the use of all of the guiding questions for each of the rubrics), her students averaged a gain of one percentile point per year on the California standards tests as compared with similar students. Students taught by a teacher at the top of the scale (44) scored, on average 20 percentile points higher than those taught by a teacher receiving the lowest passing score (24), controlling for their prior year scores and demographic characteristics. However, the sample was very small and based on a relatively unique group of candidates teaching in bilingual elementary classrooms.

Influences of Performance Assessments on Candidate and Program Learning

Other studies have looked at the influences of performance assessments on candidate learning and on feedback to programs supporting their improvement. For example, teacher education programs that participate in the PACT receive detailed, aggregated data on all of their candidates by program area and dimensions of teaching. Researchers have found that programs have used these data, as well as faculty's insights derived from scoring the portfolio assessments, to make significant changes in the curriculum sequence, individual courses, clinical experiences, and overall program design (Peck & Macdonald, 2010; Peck, Galluci, & Sloane, 2010).

Faculty and supervisors score these portfolios using standardized rubrics in moderated sessions following training, with an audit procedure to calibrate standards. They report that the scoring process causes them to reflect on their teaching and incorporate new practices that they believe will better support candidates in learning the desired skills (Darling-Hammond, 2010). Beginning teachers also report that they learn by engaging in the assessment, and evidence shows that they are later able to enact the practices they report having learned in the assessment process (Chung, 2008; Sloan, Cavazos, & Lippincott, 2007).

Similarly, studies of teachers engaging in National Board certification suggest that teachers become more conscious of their teaching decisions and change their self-reported practices as a result of this awareness and the practices required by the assessment (Athanases, 1994; Buday & Kelly, 1996; Sato, Wei & Darling-Hammond, 2008). A study of teachers' perceptions of their teaching abilities before and after completing portfolios for the National Board found that teachers reported statistically significant increases in their performance in each area assessed (planning, designing, and delivering instruction, managing the classroom, diagnosing and evaluating student learning, using subject matter knowledge, and participating in a learning community (Tracz et al., 2005). Teachers commented that videotaping their teaching and analyzing student work made them more aware of how to organize teaching and learning tasks, how to analyze student learning, and how to intervene and change course when necessary.

A survey of more than 5,600 National Board candidates found that 92% believe the National Board Certification process helped them become a better teacher, reporting that it helped them create stronger curricula, improved their abilities to evaluate student learning, and enhanced their interaction with students, parents, and other teachers (NBPTS, 2001).

In a longitudinal, quasi-experimental study that investigated learning outcomes for high school science teachers who pursued National Board Certification, Lustick and Sykes (2006) found that the certification process had a significant impact upon candidates' understanding of knowledge associated with science teaching, with a substantial overall effect size of 0.47. Teachers' knowledge was assessed before and after candidates went through the certification process through an assessment of their ability to analyze and evaluate practice.

Teachers who undertook Board certification have also been found to change their assessment practices significantly more over the course of their certification year than did teachers

who did not participate in the Board certification process (Sato, Wei, & Darling-Hammond, 2008). The most pronounced changes were in the ways teachers used a range of assessment information to support student learning.

Although there is considerable evidence that participating in performance assessments causes changes in teachers' their self-reported learning and in their practice, studies have not yet tested directly whether teachers become more effective in promoting student learning as a result of having participated in a performance assessment.

Methods

Databases and Sample

This study made use of an administrative database of California teachers who were assessed on the Performance Assessment for California Teachers (PACT) and databases that link teachers and students for three large urban school districts: Los Angeles Unified School District (LAUSD), San Diego Unified School District (SDUSD), and San Francisco Unified School District (SFUSD).

The PACT administrative database included names of 1870 candidates who completed the PACT in 2006-2008, tied to their PACT scores, plus anonymous surveys of samples of candidates who participated in PACT pilots in 2005. (Surveys were not administered in the later years.) In the surveys, candidates replied to questions about their preparation, the sources of support they received for completing the PACT assessment, and their perceptions of the educational value of the PACT for their own development as teachers. For this article, we used the surveys completed by 305 PACT candidates from eight programs who participated in PACT pilots in 2005.

The portion of the study designed to establish the predictive validity of the PACT illustrated the difficulties of tracking pre-service teachers into practice in a state without a statewide data set linking pre-service teachers to their in-service placements and linking these in-service teachers to

their students. In California, achieving these linkages required securing pre-service teachers' permissions to follow them into their school districts using name or social security number, and then finding districts that maintain linked data sets between teachers and students, which is not common in this state.

Developing a sizable sample in a few districts, even large ones, is challenging because teachers tend to disperse geographically, so that their numbers in any single district are greatly reduced compared to the total number of pre-service teachers initially studied. Furthermore, many teachers are not in tested grades or subjects and thus, once found, they do not contribute to the sample. Finally, high mobility among students in urban districts mean that it is not uncommon to have only a handful of students attached to any given teacher for a full year with tests in two consecutive years. Finally, district administrative databases often differ in the data they track, and sometimes have substantial holes or problems in their data, so combining data sets to conduct complete analyses is a challenge.

For all of these reasons, our initial sample of more than 4600 pre-service teachers with PACT scores from teacher education programs across California resulted in a final analytic sample across these three districts of 105 elementary and middle school teachers with links to students in tested grades. Specifically, of 4,622 teacher candidates who were assessed on PACT, 1,870 were later attached to names. In addition, 45 district interns were identified. Links within the three districts were then verified by comparing the subject matter of the test and the year of the PACT test compared with years of experience. ln most instances, teachers appeared as first year teachers in the year after they took the PACT (e.g., took the PACT in Spring 2006 and began teaching the 2006-07 school year). Cases were also included when their first year of teaching in the district was one year after this (e.g., taking the PACT in Spring 2006 and beginning teaching in 2007-08),

because it was reasonable that some candidates would not complete their education on schedule, or would not find a full-time position until a year later, or would switch districts after their first year. When multiple teachers had the same name, cases were only included when one matched the above criteria and the other did not.

In all, this process let to the identification of 217 LAUSD teachers, 57 SDUSD teachers, and 47 SFUSD teachers with PACT scores. Some of these teachers, however, were not assigned to tested grades and subjects, and others were not properly linked to students. Value-added models were run for grades 3-8 in ELA and 3-7 in Math (because of variability in the math tests taken by 8th grade students, as described below), so only teachers who taught students in those grades and subjects could be included. Furthermore, we also linked teacher value-added only when the subject matter of the PACT assessment (either ELA or math) matched the subject matter for the student's assessment. This led to a further winnowing of the sample. Ultimately, the ELA analysis linked students to 53 teachers in appropriate grades and subjects, and the Math analysis linked students to 52 teachers across the three districts.

Measures

PACT. Overall PACT scores are reported to candidates and programs on a scale of 1 to 4. However, there is considerable detail in the analytic scoring scheme that is used to evaluate PACT portfolios. Subscale scores are developed from scores on a set of guiding questions under each category (planning, instruction, assessment, reflection, and academic language); each of the guiding questions is also rated on a scale of 1 to 4. To take advantage of the full amount of information available, we computed subscale scores as the sum of scores for all the guiding questions within that scale, and a total score as the sum of all of the subscale scores. Thus, the highest possible total score for the PACT would be 44 points.

When teachers had more than one PACT score (which occurred when a few candidates took part both in the initial pilot conducted for test validation and also in the later finished assessment), the mean scores for all PACT assessments were used. Missing items were imputed using mean imputation, imputing the mean value for others who took the same PACT subject assessment. Data was mostly complete. Imputation was done implemented using all PACT teachers linked to these districts. Of 242 teachers so linked, at most five teachers were missing any given item.

The PACT scores for ELA and Math teachers are summarized in Table 1 (below).

Table 1: PACT Scores of Teachers Included in Analyses

PACT Scale	N	Mean	SD	Min	Max
ELA					
Total Score	53	28.30	6.78	16	43
Planning	53	8.21	2.08	4	12
Instruction	53	4.92	1.48	2	8
Assessment	53	5.03	1.52	2	8
Reflection	53	4.95	1.37	3	8
Academic Language	53	5.19	1,43	3	8
Math					
Total Score	52	27.29	5.41	18.5	39
Planning	52	8.06	1.58	5	12
Instruction	52	4.55	1.20	2	7
Assessment	52	4.92	1.43	2	8
Reflection	52	5.12	1.39	3	8
Academic Language	52	4.64	1.16	2	8

California Standards Tests (CSTs). California Standards Tests in ELA and Math are criterion-referenced tests taken yearly by all students in grades 2 through 11, except for a small number of students identified in their individualized education plan for alternative assessments. In this study, we were interested in elementary and middle school teachers.

Although we included student ELA scores for teachers who took the literacy PACT in grades 3-8, we did not include students in 8th grade math because, in contrast to lower-grade Math assessments and ELA assessments, the whole cohort of students in California do not take the same

assessment. Students in Grade 8 can take Algebra 1 or General Mathematics, depending on the course they are taking. The introduction of course selection can bias estimates if unmeasured factors influence course selection and are also correlated with achievement, so Math analyses were conducted only up through Grade 7.

In addition, a very small percentage of Grade 7 students (<1%) take Algebra 1 instead of Grade 7 Math, and these were not included in our results. Our analysis transformed all CST scores into standard scores, using data for each individual test (e.g., ELA Grade 3) to compute the mean and standard deviation.

Surveys. Brief surveys were conducted with PACT candidates during the initial pilot years of the assessment (2003-2005). Using a series of Likert scale prompts, these surveys asked candidates how well they felt their coursework and student teaching prepared them for the kind of tasks evaluated in the PACT Teaching Event, what kind of support they received in understanding how to complete the PACT from various people in their teacher education context (e.g. supervisors, instructors, cooperating teachers, fellow candidates, principals or support providers on-site), and whether they thought the Teaching Event supported their learning in several dimensions of teaching. The questions from the survey are included in the analytic tables in Appendix A. In the last year of the pilot, 305 candidates from eight programs completed the survey.

Analysis

Predictive Validity Study. We examined the influence of PACT scores on student achievement in elementary and middle school by conducting separate ordinary least squares regression analyses of students' ELA and math scores, controlling for students' prior year scores in that subject area, their demographic information, and grade level of the test. Teachers' PACT scores

were included as a predictor in these value-added models. Models were developed to include the total score and each of the PACT subscores: planning, instruction, assessment, reflection, and academic language.

Models were run for grades 3 to 8 in ELA and 3 to 7 in Math using data for 2005-06 to 2009-10. Because California Standards Tests (CSTs) are not vertically equated, and thus no scaled scores are available that have consistent meaning across tests, we standardized scores by test, subject, and grade level (e.g., Grade 4 Math). We controlled for prior year scores and student demographics, including English Learner (EL) status, sex, race/ethnicity, grade retention, and parent education.

In general, these demographic data were routinely available for students, with the exception of parent education. Within a given year, SFUSD was missing parent education for 13% of cases and LAUSD was missing parent education for 31% of cases, and SDUSD was missing parent education for 99% of cases. When parent education was missing in a given year, it was imputed, first, with the closest later year when it was available. When it was still missing, it was imputed with the most recent prior year when it was available. Using this procedure, SFUSD was missing less than 1% of cases, LAUSD was missing 29% of cases and SDUSD was still missing over 99% of cases. Because SDUSD was missing parent education because of a gap in the administrative data, an indicator variable was included to identify these cases. A separate indicator variable was included to identify cases in LAUSD and SFUSD where parent education was not known.

Finally, we controlled for district in the models. SFUSD was used as the reference group for this variable in the mathematics analysis. In the ELA model, there were no teachers with PACT literacy scores from SFUSD, so SDUSD was used as the reference group for district.

Consequential Validity Study. Descriptive analyses of the survey results first included calculation of frequency distributions, means, and standard deviations. In a set of Analyses of Variance (ANOVA), we then examined the relationships between candidates' reported agreement with statements about their learning from the Teaching Event and 1) their reports of the different levels of support they received in completing the TE; 2) their reports of coursework preparation; and 3) their reports of student teaching preparation.

Results

Relationship between PACT Scores and Teaching Effectiveness

The model explains about 69% of total variance in English language arts and about 67% in mathematics. As expected, controlling for prior year test scores, current year scores are strongly influenced by grade retention, English learner status, race / ethnicity and parent education.

In addition, teachers' overall PACT scores are significant predictors of student achievement gains in both English language arts and mathematics ($T = 3.82$; $p < .0001$ in ELA; $T=2.10$; $p < .05$ in mathematics). (See Tables A-1 and A-2, in the appendix.) In addition, some of the PACT sub-scores are associated with value-added achievement gains. In ELA, the subscales that are significant predictors of student achievement include Planning ($p < 0.01$), Assessment ($p < 0.01$)), and Academic Language ($p < 0.01$). In Mathematics, the subscales that are significant predictors of achievement include Assessment ($p <0.01$), and Reflection ($p <0.01$).

In separate analyses separating elementary and middle school teachers, we found that PACT's predictive power was particularly strong in the middle school grades, where PACT scores were even more significant influences on student achievement. The strength of the relationship was only slightly stronger for English language arts ($T= 4.03$; $p < .0001$), but it was noticeably stronger in mathematics ($T= 6.37$; $p < 0.0001$).

In order to describe the magnitude of the relationships, we estimated an effect size by contrasting predicted achievement for students of PACT teachers with the lowest observed PACT scores (barely passing) versus the predicted achievement of identical students with PACT teachers having the highest observed PACT scores. On average, in ELA, this effect size was .15 for PACT Total Score, .18 for PACT Planning, .14 for PACT Assessment, and .13 for PACT Academic Language. In Math, the effect sizes were .11 for PACT Total Score, .17 for Assessment, and .17 for Reflection. Larger effect sizes were observed applying the same procedure in Middle School. In ELA Middle School, the effect sizes were .17 for Total PACT score, .14 for Planning, .17 for Assessment, and .18 for Academic Language. In Math Middle School, the effect sizes were .43 for PACT Total Score, .31 for Planning, .22 for Instruction, .45 for Assessment, and .56 for Reflection.

Relationship between PACT Completion and Teacher Learning

As part of the pilot studies, student teachers were asked a set of questions about whether they felt they learned more about teaching by completing the PACT. They were also asked how well their programs prepared them to succeed on the PACT -- in terms of their coursework and student teaching experiences, as well as direct support for understanding how to assemble the portfolio.

A substantial majority of student teachers agreed that constructing the teaching event was a source of learning for them and that they expected it to be useful for their future teaching practice. On specific questions, 67.2% agreed that they learned important skills through the process of constructing the teaching event; 63.6% agreed that it helped them improve their knowledge of learners; 68.6% agreed that it helped improve their assessment of student learning progress; and 70.5% agreed that constructing the TE helped them to reflect more carefully on instructional decisions. (See table A-3.)

Large majorities of candidates also agreed that their teacher education experiences were helpful in preparing them for the PACT teaching event. Overall, 83.6% agreed that their coursework helped prepare them for the TE, and 86.3% agreed that their student teaching experiences helped prepare them for the TE. They rated their university supervisors and other credential candidates as the most helpful resources in supporting them in completing the TE. (See tables A-4 and A-5.)

These ratings of the quality of preparation and the help received in preparing the TE were, in turn, strongly and significantly related to candidates' perceptions that they learned from completing the teaching event. (See tables A-6 through A-8.) Candidates who rated their coursework and student teaching to be useful preparation were much more likely to say they found the experience of conducting the TE educative on each dimension of learning ($p < .0001$): learning important skills, improving lesson planning, improving knowledge of learners, improving assessment of student learning progress, improving implementation of instruction, and improving reflection on instructional decisions. They were also more likely to say that the TE enhanced their teacher preparation experience and would be useful in their future teaching ($p < .0001$). Similarly, those who felt better supported in the course of completing the TE also found it more helpful to their own learning and future teaching ($p < .0001$ on each indicator).

Discussion

The Performance Assessment for California Teachers was developed by teacher educators and K-12 teachers to provide a valid measure for evaluating the readiness of prospective teachers to teach based on actual evidence of their abilities to plan, implement instruction, assess learning, and adapt their teaching to the needs of learners. In addition, PACT developers sought to create an

assessment that would support the learning of prospective teachers and provide useful information to teacher education faculty about how to improve their programs.

In addition to research establishing the reliability and validity of PACT based on traditional psychometric analyses (Pecheone & Chung, 2006, 2007), the studies reported here examined the predictive validity of PACT Teaching Event scores (that is, whether they are related to candidates' later teaching effectiveness) and the consequential validity of the assessment (that is, what the consequences of engaging in the assessment are for candidates, themselves). These are aspects of test validity that are rarely examined, but that are essential for understanding the educational value of assessments (Messick, 1989; 1994).

It is important to note that, as with all licensing tests, studies evaluating relationships between test performance and later effectiveness suffer from a truncated range of observed scores, as those candidates who did not pass the test are not licensed to practice, and hence do not appear in the later sample. As a consequence, we cannot estimate the differential effects of a candidate whose performance was too poor to pass the assessment in comparison with candidates who met the passing standard.

With respect to predictive validity, even within this truncated range, we found that candidates' PACT scores on the elementary literacy and mathematics assessments were positively and significantly related to their later teaching effectiveness, as measured by their students' performance on California standards tests in English language arts and mathematics, respectively. These analyses controlled for students' prior achievement and demographic characteristics, as well as district and grade level of the tests.

Certain subscale scores on the PACT, measuring dimensions of teaching, were also predictive of student gains. In ELA, the planning, assessment, and academic language subscales

were significant predictors of teaching effectiveness. In mathematics, the assessment and reflection subscales were significant.

It is perhaps not surprising that the assessment component of the PACT is a predictor of teachers' effectiveness for both ELA and mathematics, since it calls on teachers to examine students' learning of the curriculum they are teaching, analyze student work for the class to discover what is being learned by whom, and where struggles are occurring, and plan for additional teaching to address these student needs. This is perhaps one of the most important marks of an effective teacher -- one who teaches *in relation to* student learning, rather than merely by covering the curriculum, irrespective of student responses.

It is also not surprising that the planning dimension may be more important in ELA than in math, given the fact that planning in mathematics is more often guided by a textbook sequence, whereas planning for reading and writing instruction often draws on a wider array of materials and teaching strategies. It is interesting that reflection might prove especially important in mathematics - where teachers' ability to think carefully about *how* they would teach or re-teach the content is a possible contributor to effectiveness.

Finally, the importance for literacy gains of teachers' capacities to support academic language development is logical. This dimension of the assessment looks both at how teachers support access to the content they are teaching for new English language learners and at how they explicitly plan for and teach the academic language of the discipline to all of their learners. This has proved to be an important area of learning both for candidates and for their programs, as the knowledge base for developing students' academic language was not widely taught when the PACT was first introduced.

We found that the strength of PACT's influence on later effectiveness was larger for the subsample of middle school teachers than for the sample as a whole, and this was especially true in mathematics. We wonder whether the subject-specific nature of the PACT assessment is particularly important as content demands increase in the upper grades and as teaching is more departmentalized. Content pedagogical skills may be particularly salient in mathematics, where teachers' content knowledge and content pedagogical knowledge are highly variable. PACT may be capturing some of these aspects of mathematics knowledge for teaching in the upper grades.

The relationships between PACT scores and the later achievement of candidates' students, though statistically significant, are relatively small. In addition to the truncated range of scores associated with licensing tests, there are many mediating variables that are unmeasured in this study. These include the conditions associated with beginning teachers' initial teaching assignments -- for example, class sizes, curriculum materials, quality and collaboration among teaching peers, and the organization of teaching supports -- and the kind and quality of mentoring beginning teachers may have received. Because of limitations in the data bases, we were also unable to measure the effects of classroom composition on student outcomes, which are known to be substantial. Future research should seek to examine these context variables.

Finally, we found that candidates felt they learned from engaging in the PACT assessment. Large majorities felt that they learned important skills through the process of constructing the teaching event, and that the experience helped them improve their knowledge of learners, their assessment of student learning progress, and their ability to reflect more carefully on instructional decisions. Well over 80 percent of candidates felt that their teacher education coursework and student teaching experiences were helpful in preparing them for the PACT teaching event.

Interestingly, the more candidates felt well-supported by their programs in learning to teach and in completing the assessment, the more they felt they learned from the assessment experience. Thus, the consequential validity of the assessment appeared to be strengthened as it was embedded in a supportive environment for learning to teach.

There is certainly considerable work to be done to fine-tune these kinds of performance assessments and to evaluate the ways in which they may be most productively used. We conclude from this study that this kind of work is worth continuing: It appears possible to construct authentic evaluation tools for beginning teachers that are valid indicators of their readiness to teach and that may, in addition, contribute to their process of learning to teach.

Table A-1

Relationship between PACT Scores / Subscores and ELA Value-Added Grades 3-8

Parameter	Total Score	Planning	Instruction	Assessment	Reflection	Academic Language
PACT Score/Subscore	0.005 (0.001)**	0.023 (0.005)**	0.010 (0.006)	0.023 (0.005)**	0.008 (0.007)	0.026 (0.007)**
Prior ELA	0.75 (0.01)**	0.75 (0.01)**	0.75 (0.01)**	0.75 (0.01)**	0.75 (0.01)**	0.75 (0.01)**
English Learner	-0.18 (0.03)**	-0.18 (0.03)**	-0.18 (0.03)**	-0.18 (0.03)**	-0.18 (0.03)**	-0.18 (0.03)**
Redesignated Fluent	0.00 (0.02)	0.01 (0.02)	0.01 (0.02)	0.00 (0.02)	0.01 (0.02)	0.00 (0.02)
Female	0.06 (0.02)**	0.06 (0.02)**	0.06 (0.02)**	0.07 (0.02)**	0.06 (0.02)**	0.07 (0.02)**
Asian	0.17 (0.06)**	0.16 (0.06)	0.17 (0.06)**	0.18 (0.06)	0.16 (0.06)**	0.17 (0.06)
African-American	-0.13 (0.03)**	-0.14 (0.03)**	-0.14 (0.03)**	-0.14 (0.03)**	-0.14 (0.03)**	-0.14 (0.03)**
Caucasian	0.17 (0.07)*	0.17 (0.07)*	0.16 (0.07)*	0.17 (0.07)*	0.16 (0.07)*	0.17 (0.07)*
Filipino	-0.03 (0.1)	0.00 (0.1)	-0.02 (0.1)	-0.02 (0.1)	-0.02 (0.10)	-0.03 (0.1)
Native American	-0.18 (0.17)	-0.18 (0.17)	-0.17 (0.17)	-0.17 (0.17)	-0.17 (0.17)	-0.17 (0.17)
Pacific Islander	0.49 (0.25)*	0.50 (0.25)	0.49 (0.25)*	0.50 (0.25)*	0.49 (0.25)*	0.47 (0.25)
Parent Education						
Graduate School	0.16 (0.06)**	0.16 (0.06)**	0.16 (0.06)**	0.16 (0.06)*	0.16 (0.06)*	0.16 (0.06)*
College Graduate	0.01 (0.04)	0.01 (0.04)	0.01 (0.04)	0.01 (0.04)	0.01 (0.04)	0.01 (0.04)
Some College	0.02 (0.03)	0.02 (0.03)	0.02 (0.03)	0.02 (0.03)	0.02 (0.03)	0.02 (0.03)
Not HS Grad	-0.03 (0.02)	-0.03 (0.02)	-0.03 (0.02)	-0.03 (0.02)	-0.03 (0.02)	-0.03 (0.02)
Missing	0.00 (0.02)	0.00 (0.02)	0.00 (0.02)	0.00 (0.02)	0.00 (0.02)	0.00 (0.02)
Grade 3	-0.08 (0.04)*	-0.09 (0.04)*	-0.08 (0.04)*	-0.09 (0.04)*	-0.08 (0.04)*	-0.08 (0.04)*
Grade 4	-0.06 (0.04)	-0.07 (0.04)	-0.05 (0.04)	-0.07 (0.04)	-0.05 (0.04)	-0.05 (0.04)
Grade 6	-0.02 (0.03)	-0.02 (0.03)	-0.01 (0.03)	-0.03 (0.03)	-0.01 (0.03)	-0.02 (0.03)
Grade 7	0.04 (0.02)	0.03 (0.02)	0.05 (0.02)*	0.03 (0.02)	0.06 (0.02)*	0.04 (0.02)
Grade 8	0.03 (0.03)	0.04 (0.03)	0.06 (0.03)	0.04 (0.03)	0.06 (0.03)*	0.05 (0.03)
Retained	0.32 (0.16)	0.32 (0.16)*	0.33 (0.16)*	0.31 (0.16)	0.33 (0.17)*	0.32 (0.17)
LAUSD	-0.05 (0.05)	-0.04 (0.05)	-0.09 (0.05)	-0.08 (0.05)	-0.11 (0.05)*	-0.06 (0.05)
Constant	-0.04 (0.08)	-0.08 (0.08)	0.09 (0.07)	0.03 (0.06)	0.11 (0.07)	-0.01 (0.07)
R-Squared (Adjusted)	0.689	0.689	0.688	0.689	0.688	0.689
Teachers (N)	53	53	53	53	53	53
Students (N)	4060	4060	4060	4060	4060	4060

Table A-2

Relationship between PACT Scores / Subscores and Math Value-Added Grades 3-7

Parameter	Total	Planning	Instruction	Assessment	Reflection	Academic Language
PACT Score/Subscore	0.005 (0.002)*	0.013 (0.008)	-0.017 (0.011)	0.029 (0.009)**	0.034 (0.010)**	0.007 (0.011)
Prior Math	0.73 (0.01)**	0.73 (0.01)**	0.73 (0.01)**	0.73 (0.01)**	0.73 (0.01)**	0.73 (0.01)**
English Learner	-0.14 (0.03)**	-0.14 (0.03)**	-0.14 (0.03)**	-0.14 (0.03)**	-0.13 (0.03)**	-0.14 (0.03)**
Redesignated Fluent	0.10 (0.03)**	0.10 (0.03)	0.09 (0.03)	0.10 (0.03)**	0.10 (0.03)**	0.10 (0.03)
Female	-0.01 (0.02)	0.00 (0.02)**	0.00 (0.02)**	-0.01 (0.02)	-0.01 (0.02)	0.00 (0.02)
Asian	0.33 (0.05)**	0.33 (0.05)**	0.34 (0.05)**	0.33 (0.05)**	0.33 (0.05)**	0.33 (0.05)**
African-American	-0.14 (0.04)**	-0.14 (0.04)	-0.14 (0.04)	-0.15 (0.04)**	-0.14 (0.04)**	-0.14 (0.04)**
Caucasian	0.10 (0.06)	0.10 (0.06)*	0.13 (0.06)*	0.10 (0.06)	0.11 (0.06)	0.11 (0.06)
Filipino	0.08 (0.08)	0.08 (0.08)	0.09 (0.08)	0.07 (0.08)	0.09 (0.08)	0.09 (0.08)
Native American	-0.24 (0.17)	-0.25 (0.17)	-0.24 (0.17)	-0.24 (0.17)	-0.24 (0.17)	-0.24 (0.17)
Pacific Islander	0.03 (0.13)	0.02 (0.13)	0.03 (0.13)	0.02 (0.13)	0.03 (0.13)	0.03 (0.13)
Other Ethnicity	0.18 (0.07)	0.18 (0.07)**	0.19 (0.07)**	0.18 (0.07)*	0.19 (0.07)**	0.19 (0.07)*
Parent Education						
Graduate School	0.18 (0.06)	0.18 (0.06)	0.18 (0.06)	0.18 (0.06)**	0.19 (0.06)**	0.18 (0.06)**
College Graduate	0.04 (0.04)	0.05 (0.04)	0.04 (0.04)	0.04 (0.04)	0.05 (0.04)	0.04 (0.04)
Some College	0.03 (0.04)	0.03 (0.04)	0.03 (0.04)	0.03 (0.04)	0.04 (0.04)	0.03 (0.04)
Not HS Grad	-0.01 (0.03)	-0.01 (0.03)	-0.01 (0.03)	-0.01 (0.03)	-0.01 (0.03)	-0.01 (0.03)
Missing	-0.02 (0.03)	-0.02 (0.03)**	-0.02 (0.03)	-0.01 (0.03)	-0.02 (0.03)	-0.02 (0.03)
Grade 3	-0.20 (0.04)**	-0.20 (0.04)	-0.20 (0.04)**	-0.20 (0.04)**	-0.21 (0.04)**	-0.20 (0.04)**
Grade 4	0.05 (0.03)	0.05 (0.03)**	0.06 (0.03)	0.04 (0.03)	0.07 (0.03)	0.05 (0.03)
Grade 6	-0.18 (0.04)**	-0.18 (0.04)*	-0.18 (0.04)**	-0.15 (0.04)**	-0.18 (0.04)**	-0.18 (0.04)**
Grade 7	0.08 (0.03)*	0.08 (0.03)	0.05 (0.03)	0.09 (0.03)**	0.08 (0.03)**	0.06 (0.04)
Retained	0.36 (0.2)	0.36 (0.2)	0.35 (0.2)	0.37 (0.20)	0.35 (0.2)	0.36 (0.2)
LAUSD	-0.11 (0.04)**	-0.11 (0.04)	-0.1 (0.04)*	-0.13 (0.04)**	-0.11 (0.04)**	-0.10 (0.04)**
SDUSD	-0.17 (0.14)	-0.16 (0.14)	-0.09 (0.14)	-0.18 (0.14)	-0.14 (0.14)	-0.13 (0.14)
Constant	0.02 (0.08)	0.04 (0.08)	0.21 (0.07)**	0.02 (0.07)	-0.03 (0.07)	0.11 (0.07)
R-Squared (Adjusted)	0.668	0.668	0.668	0.669	0.669	0.668
Teachers (N)	52	52	52	52	52	52
Students (N)	3196	3196	3196	3196	3196	3196

Notes: * p < .05; ** p < .01
Numbers in parentheses are standard errors.

Table A-3. Candidates' Perspectives on the PACT Teaching Event

Indicate your level of agreement	N	Missing	Strongly Disagree =1	Disagree =2	Agree =3	Strongly Agree =4	Mean	Std. Dev.
Please indicate your level of agreement with each of the statements about the PACT Teaching Event below:								
1. I learned important skills through the process of constructing the Teaching Event	305	4 1.3%	35 11.5%	61 20.0%	159 52.1%	46 15.1%	2.72	.862
2. The process of constructing the teaching Event helped to improve my lesson planning	305	5 1.6%	40 13.1%	93 30.5%	123 40.3%	44 14.4%	2.57	.899
3. The process of constructing the Teaching Event helped to improve my knowledge of learners	305	5 1.6%	35 11.5%	71 23.3%	148 48.5%	46 15.1%	2.68	.871
4. The process of constructing the Teaching Event helped to improve my assessment of student learning progress	305	5 1.6%	32 10.5%	59 19.3%	149 48.9%	60 19.7%	2.79	.884
5. The process of constructing the Teaching Event helped me to improve my implementation of instruction	305	4 1.3%	37 12.1%	65 21.3%	149 48.9%	50 16.4%	2.70	.888
6. The process of constructing the Teaching Event helped me to reflect more carefully on my instructional decisions	305	5 1.6%	31 10.2%	51 16.7%	137 44.9%	81 26.6%	2.89	.919
7. My teacher preparation experience was enhanced by the Teaching Event	305	4 1.3%	51 16.7%	81 26.6%	126 41.3%	43 14.1%	2.53	.936
8. The Teaching Event will be useful for my future teaching practice	305	4 1.3%	50 16.4%	70 23.0%	132 43.3%	49 16.1%	2.60	.949

Table A-4. Candidates' Perspectives on Support for Completing the Teaching Event

How helpful have the following people been as you completed your Teaching Event?	N	N/A	Not Very Helpful =1	2	3	4	Very Helpful = 5	Mean	Std. Dev.
1. Credential candidates also completing the Teaching Events	302	22 7.3%	14 4.6%	18 6.0%	54 17.9%	88 29.1%	106 35.1%	3.62	1.489
2. My university supervisor	301	17 5.6%	23 7.6%	24 8.0%	55 18.3%	83 27.6%	99 32.9%	3.53	1.487
3. My cooperating/master teacher	302	70 23.2%	19 6.3%	24 7.9%	58 19.2%	61 20.2%	70 23.2%	2.76	1.872
4. University instructors/professors	301	49 16.3%	23 7.6%	32 10.6%	74 24.6%	64 21.3%	59 19.6%	2.86	1.692
5. School administrators/support providers	300	101 33.7%	60 20.0%	45 15.0%	40 13.3%	28 9.3%	26 8.7%	1.71	1.660

Table A-5. Candidates' Perspectives on Preparation to Complete the Teaching Event and Preparation to Teach

Indicate your level of agreement	N	Missing	Strongly Disagree =1	Disagree =2	Agree =3	Strongly Agree =4	Mean	Std. Dev.
1. My teacher preparation courses have helped prepare me to complete my teaching event.	305	6 2.0%	12 3.9%	32 10.5%	141 46.2%	114 37.4%	3.19	783
2. My student/intern teaching placement experiences have helped prepare me to complete my teaching event	305	6 2.0%	4 1.3%	32 10.5%	128 42.0%	135 44.3%	3.32	.716
3. My teacher preparation courses have helped prepare me to teach independently in my own class.	305	6 2.0%	9 3.0%	32 10.5%	139 45.6%	119 39.0%	3.23	.758
4. My student/intern teaching placement experiences have helped prepare me to teach independently in my own class.	305	4 1.3%	2 .7%	12 3.9%	87 28.5%	200 65.6%	3.61	.599
5. I am confident of my ability to teach all students to high levels.	305	5 1.6%	3 1.0%	27 8.9%	121 39.7%	149 48.9%	3.39	.692

Table A-6. ANOVA – Differences in Reported Candidate Learning with Different Levels of Support

Survey item		Sum of Squares	df	Mean Square	F	Sig.
TE helped me learn important skills	Between Groups	22.686	4	5.671	8.487	.000
	Within Groups	195.119	292	.668		
	Total	217.805	296			
TE helped improve lesson planning	Between Groups	23.469	4	5.867	7.951	.000
	Within Groups	214.744	291	.738		
	Total	238.213	295			
TE helped improve knowledge of learners	Between Groups	28.524	4	7.131	10.648	.000
	Within Groups	194.882	291	.670		
	Total	223.405	295			
TE helped improve assessment of student learning progress	Between Groups	26.511	4	6.628	9.489	.000
	Within Groups	203.960	292	.698		
	Total	230.471	296			
TE helped improve implementation of instruction	Between Groups	21.011	4	5.253	7.272	.000
	Within Groups	210.915	292	.722		
	Total	231.926	296			
TE helped improve my reflection on instructional decisions	Between Groups	21.506	4	5.376	6.953	.000
	Within Groups	225.035	291	.773		
	Total	246.541	295			
TE helped enhance teacher preparation experience	Between Groups	28.949	4	7.237	9.319	.000
	Within Groups	226.775	292	.777		
	Total	255.724	296			
TE useful for future teaching practice	Between Groups	29.039	4	7.260	8.972	.000
	Within Groups	236.281	292	.809		
	Total	265.320	296			

Note: Independent Variable: Total Support Score (for completing the TE) is calculated from the sum of responses to the question, "How helpful have the following people been as you completed your Teaching Event?" where each of the following categories is rated from 1-5 (Not Very Helpful =1 ...Very Helpful = 5): Credential candidates also completing the Teaching Events, My university supervisor, My cooperating/master teacher, University instructors/professors, School administrators/ support providers. From these sum of these responses, the Total Support Score (for completing the TE) is recorded as follows: 1=1-5, 2=6-10, 3=11-15, 4=16-20, 5=21-25

Table A-7. ANOVA – Differences in Reported Candidate Learning by Coursework Preparation

Independent Variable: "My teacher preparation courses have helped prepare me to complete my teaching event"
(1=Strongly Disagree, 2=Disagree, 3=Agree, 4=Strongly Agree)

Survey item		Sum of Squares	df	Mean Square	F	Sig.
TE helped me learn important skills	Between Groups	60.919	3	20.306	37.928	.000
	Within Groups	157.403	294	.535		
	Total	218.322	297			
TE helped improve lesson planning	Between Groups	55.061	3	18.354	29.309	.000
	Within Groups	183.484	293	.626		
	Total	238.545	296			
TE helped improve knowledge of learners	Between Groups	56.385	3	18.795	32.953	.000
	Within Groups	167.117	293	.570		
	Total	223.502	296			
TE helped improve assessment of student learning progress	Between Groups	69.427	3	23.142	42.105	.000
	Within Groups	161.044	293	.550		
	Total	230.471	296			
TE helped improve implementation of instruction	Between Groups	53.454	3	17.818	29.406	.000
	Within Groups	178.146	294	.606		
	Total	231.601	297			
TE helped improve my reflection on instructional decisions	Between Groups	52.043	3	17.348	26.076	.000
	Within Groups	194.926	293	.665		
	Total	246.970	296			
TE helped enhance teacher preparation experience	Between Groups	71.441	3	23.814	37.560	.000
	Within Groups	186.401	294	.634		
	Total	257.842	297			
TE useful for future teaching practice	Between Groups	62.168	3	20.723	30.027	.000
	Within Groups	202.896	294	.690		
	Total	265.064	297			

Table A-8. ANOVA – Differences in Reported Candidate Learning by Student Teaching Preparation

Independent Variable: "My student/intern teaching placement experiences have helped prepare me to complete my teaching event"
(1=Strongly Disagree, 2=Disagree, 3=Agree, 4=Strongly Agree)

Survey item		Sum of Squares	df	Mean Square	F	Sig.
TE helped me learn important skills	Between Groups	37.911	3	12.637	20.449	.000
	Within Groups	181.689	294	.618		
	Total	219.601	297			
TE helped improve lesson planning	Between Groups	36.730	3	12.243	17.738	.000
	Within Groups	202.239	293	.690		
	Total	238.970	296			
TE helped improve knowledge of learners	Between Groups	27.178	3	9.059	13.444	.000
	Within Groups	197.435	293	.674		
	Total	224.613	296			
TE helped improve assessment of student learning progress	Between Groups	33.680	3	11.227	16.569	.000
	Within Groups	198.529	293	.678		
	Total	232.209	296			
TE helped improve implementation of instruction	Between Groups	30.361	3	10.120	14.872	.000
	Within Groups	200.059	294	.680		
	Total	230.419	297			
TE helped improve my reflection on instructional decisions	Between Groups	26.098	3	8.699	11.418	.000
	Within Groups	223.235	293	.762		
	Total	249.333	296			
TE helped enhance teacher preparation experience	Between Groups	36.205	3	12.068	15.991	.000
	Within Groups	221.889	294	.755		
	Total	258.094	297			
TE useful for future teaching practice	Between Groups	33.914	3	11.305	14.206	.000
	Within Groups	233.956	294	.796		
	Total	267.869	297			

Acknowledgements

The research reported in this article was funded by a generous grant from the Carnegie Corporation of New York. The authors would like to thank individuals in the PACT consortium who assisted in the facilitation of the data collection process, and in the research divisions of the three school districts that provided teacher and student data: Los Angeles, San Diego, and San Francisco Unified School Districts. We would also like to thank colleagues Kendyll Stansbury and Ray Pecheone, who assisted in securing and organizing data for the analyses.

References

Amrein-Beardsley, A. (2008). Methodological concerns about the education value-Added assessment system. *Educational Researcher, 37(2)*, 65-75.

Andrews, J.W., Blackmon, C.R., & Mackey, J.A. (1980). Preservice performance and the National Teacher Examinations. *Phi Delta Kappan, 61(5)*: 358-359.

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

Ayers, J.B., & Qualls, G.S. (1979). Concurrent and predictive validity of the National Teacher Examinations. *Journal of Educational Research, 73 (2)*: 86-92.

Buday, M., & Kelly, J. (1996). National Board certification and the teaching professions commitment to quality assurance. *Phi Delta Kappan, 78*, 215-219.

Campbell, J., Kyriakides, L, Muijs, D, and Robinson, W. (2004). *Assessing teacher effectiveness: Developing a differentiated model*. New York, NY: RoutledgeFalmer.

Cantrell, S., Fullerton, J., Kane, T. J., and Staiger., D. O. (2008). National board certification and teacher effectiveness: Evidence from a random assignment experiment. NBER Working Paper 14608. Downloaded 19 May 2009 from http://www.nber.org/papers/w14608.

Cavaluzzo, L. (2004). *Is National Board Certification an effective signal of teacher quality?* (National Science Foundation No. REC-0107014). Alexandria, VA: The CNA Corporation.

Chung, R. R. (2008). Beyond assessment: Performance assessments in teacher education. *Teacher Education Quarterly, 35* (1).

Clotfelter, C., Ladd, H., & Vigdor, J. (2006). *Teacher-student matching and the assessment of teacher effectiveness* (NBER Working paper 11936). Cambridge, MA: National Bureau of Economic Research.

Clotfelter, C., Ladd, H., & Vigdor, J. (2007). *How and why do teacher credentials matter for student achievement?* (NBER Working Paper 12828). Cambridge, MA: National Bureau of Economic Research.

Darling-Hammond, L. (2007). *Recognizing and Enhancing Teacher Effectiveness: A Policy Maker's Guide*. Washington, DC: Council for Chief State School Officers.

Darling-Hammond, L. (2010). *Evaluating Teacher Effectiveness: How Teacher Performance Assessments can Measure and Improve Teaching.* Washington DC: Center for American Progress.

Darling-Hammond, L.. Chung, R., and Frelow, F. (2002). Variation in teacher preparation: How well do different pathways prepare teachers to teach? *Journal of Teacher Education, 53*(4), 286-302 .

Floden, R., and Meniketti, M. (2005). Research on the effects of coursework in the arts and sciences and the foundations of education. In M. Cochran-Smith and K. M. Zeichner (Eds.) *Studying teacher education: The report of the AERA panel on research and teacher education*. Mahweh, NJ: Lawrence Erlbaum Associates, Inc.

Goldhaber, D., and Anthony, E. (2007). Can teacher quality be effectively assessed? National board certification as a signal of effective teaching. *The Review of Economics and Statistics, 89*(1), 134-150.

Haney, W., Madaus, G., & Kreitzer, A. (1987). Charms talismanic: testing teachers for the improvement of American education. In E.Z. Rothkopf (Ed.) *Review of Research in Education, 14*: 169-238. Washington, DC: American Educational Research Association.

Harris, D., and Sass, T.R. (2007). *The effects of NBPTS-certified teachers on student achievement.* Downloaded 10 October, 2010 from http://www.nbpts.org/userfiles/file/harris_sass_final_2007.pdf.

Lustick, D., & Sykes, G. (2006). National Board Certification as professional development: What are teachers learning? *Education Policy Analysis Archives, 14*(5). Retrieved June 10, 2008, from http://epaa.asu.edu/epaa/v14n5/

Messick, S. (1989). Validity. In R.L. Linn (Ed.), *Educational measurement* (3rd Ed.) (pp. 13-103). New York: American Council on Education/Macmillan.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher, 23*(2), 13-23.

Mitchell, K. J., Robinson, D. Z., Plake, B. S., & Knowles, K. T. (Eds.). (2001). *Testing teacher candidates: The role of licensure tests in improving teacher quality*. Washington, DC: National Academy Press.

National Board for Professional Teaching Standards (NBPTS). (2001). *"I am a better teacher:" What candidates for National Board certification say about the assessment process.* Arlington, VA: Author.

Newton, S. P. (2010). *Predictive validity of the performance assessment for California teachers*. Stanford, CA: Stanford Center for Opportunity Policy in Education. Retrieved from http://scale.stanford.edu/.

Pecheone, R. L.., and Chung, R. R. (2006). Evidence in teacher education: The performance assessment for California teachers. *Journal of Teacher*; 57(1), 22-36.

Pecheone, R. & Chung, R. (2007). *Technical Report of the Performance Assessment for California Teachers (PACT): Summary of Validity and Reliability Studies for the 2003-04 Pilot Year.* Stanford, CA: Stanford University.

Peck, C., Galluci, C., and Sloan, T., (2010). Negotiating Implementation of High-Stakes Performance Assessment Policies in Teacher Education: From Compliance to Inquiry. *Journal of Teacher Education*, *61* (5); 451-463.

Peck, C. & McDonald, M., (2010, January). *Creating "Cultures of Evidence" in Teacher Education:Context, Policy and Practice in Three High Data Use Programs*. Report submitted to the Agency for Education Development.

Rubin, D. B., Stuart, E. A., and Zanutto, E. L. (2004). A potential outcomes view of value-added assessment in education. *Journal of Educational and Behavioral Statistics*, *29*(1), 103-116.

Sanders, W. L., Ashton, J. J., & Wright, S. P. (2005). Comparison of the effects of NBPTS [National Board for Professional Teaching Standards] certified teachers with other teachers on the rate of student academic progress. Arlington, VA: National Board for Professional Teaching Standards. Downloaded 19 May 2009 from: http://www.nbpts.org/UserFiles/File/SAS_final_NBPTS_report_D_-_Sanders.pdf

Sato, M., Chung Wei, R., & Darling-Hammond, L. (2008). Improving teachers' assessment practices through professional development: The case of National Board Certification. *American Educational Research Journal, 45*: 669-7000.

Sloan, T., Cavazos, L., & Lippincott, A. (2007, April). *A holistic approach to assessing teacher competency: Can one assessment do it all?* Paper presented at the Annual Meeting of the American Educational Research Association, Chicago.

Smith, T., Gordon, B., Colby, S., & Wang, J. (2005). *An examination of the relationship between the depth of student learning and National Board certification status*. Boone, NC: Appalachian State University, Office for Research on Teaching.

Tracz, S. M., Daughtry, J., Henderson-Sparks, J., Newman, C., Sienty, S. (2005). The Impact of NBPTS Participation on Teacher Practice: Learning from Teacher Perspectives. *Educational Research Quarterly, 28* (3): 35-50.

Vandevoort, L. G., Amrein-Beardsley, A. & Berliner, D. C. (2004). National Board Certified teachers and their students' achievement. *Education Policy Analysis Archives, 12* (46), pp. 1-117.

Wilson, M., Hallam, P. J., Pecheone, R., and Moss, P. (2007). Using Student Achievement Test scores as evidence of external validity for indicators of teacher quality: Connecticut's Beginning Educator Support and Training program. Berkeley, CA: Berkeley Evaluation and Research Center.