

**Preservice Performance Assessment and  
Teacher Early Career Effectiveness:  
Preliminary Findings on the  
Performance Assessment for California Teachers**

**Stephen Newton  
Stanford University**

This study was conducted by the Stanford Center for Assessment, Learning, and Equity,  
with support from the Carnegie Corporation of New York and  
the Morgan Family Foundation.

© 2010 Stanford Center for Assessment, Learning, and Equity. All rights reserved.

Citation: Newton, S. (2010). *Preservice performance assessment and teacher early career effectiveness: Preliminary findings on the performance assessment for California teachers*. Stanford, CA: Stanford University, Stanford Center for Assessment, Learning, and Equity.

**Stanford Center for Assessment, Learning, and Equity**  
Barnum Center, 505 Lasuen Mall  
Stanford, California 94305  
Phone: 650.725.0703  
<http://scale.stanford.edu/>

## Abstract

This study presents preliminary findings on the relationship between beginning teacher's scores on the Performance Assessment for California Teachers (PACT), a preservice performance assessment of a wide range of teaching skills, and their subsequent teaching effectiveness measured by students' value-added achievement gains in English Language Arts (ELA). Linking PACT scores and four separate value-added estimates for 14 first- and second-year teachers with 259 students in third through sixth grade in an urban school district, total PACT score correlated approximately .50 with teacher value-added (ranging from .46 to .53, depending on the model), and subscores for the assessment dimension of the portfolio (evaluating candidates' ability to use assessment data to support student learning) correlated .58 to .66 with value-added. PACT scores predicted very substantial differences in teacher value-added. For each additional point a teacher scored on PACT (evaluated on a 44-point scale), her students averaged a gain of one percentile point per year on the California Standards Tests as compared with similar students. Thus, 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). In this limited sample, PACT was a strong predictor of teacher effectiveness measured by students' value-added achievement gains. Future studies will seek to increase the sample sizes with more complete data.



# Introduction

The Performance Assessment for California Teachers (PACT) has been developed to measure readiness of beginning teachers to teach by assessing them in a variety of teaching proficiencies, demonstrated in actual classrooms (Pecheone & Chung, 2006). 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. The present study, linking performance on the PACT of preservice teachers with their early-career effectiveness measured by value-added assessment of their effectiveness, addresses the important issue of the criterion-validity of the assessments.

The study was conducted as a follow-up to pilot tests of the assessment in 2005 to 2008, tracking the later value-added scores of students of a small cohort of multiple-subject teachers in San Diego during their first two years in the classroom. 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). A larger-scale version of this small exploratory study is currently underway, using data from multiple cities in California.

## Rationale for the Study

The ability of performance assessments to predict 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 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 were these scores used for later evaluation). As Darling-Hammond (2010) notes:

Since most experts agree that at least three years of data about a given teacher are necessary to achieve a modicum of stability, the direct use of student test score data to evaluate teachers does not help inform judgments about new entrants to the profession. Yet, in order to protect students, governments must make judgments about whether professionals are well enough prepared to practice safely and competently as soon as they enter the profession.

Finally, 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, through methods such as value-added analysis, because of their timing. Preservice performance assessments provide quicker information to teacher education programs because it takes a year or more for administrative databases of student achievement to be available, if at all. Furthermore, once beginning teachers leave their preparation institutions, they can be affected positively or negatively by conditions at their school sites, so it is difficult to tease apart the effects of preservice education from other sources of influence (Darling-Hammond, Chung, & Frelow, 2005; Floden & Meniketti, 2005). This institutional feedback has particular usefulness because such assessments given within a program provide quicker feedback to the teacher education program itself that can help inform it about areas of strength and weakness.

## **The Performance Assessment for California Teachers**

The Performance Assessment for California Teachers (PACT) has been developed as an instrument for measuring teaching knowledge, that is, knowledge as actually applied in a teaching situation. This information serves as a criterion for making credentialing decisions for teachers, requiring that they demonstrate a breadth of proficiencies relevant to teaching effectiveness (Pecheone & Chung, 2006). It also has 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. 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).

PACT consists of two classes of assessments, embedded signature assessments to be completed throughout the year, and a summative assessment of teaching knowledge and skills during student teaching, also known as the teaching event or TE (Pecheone and 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)

Candidate work is then rated on a number of elements defined by subject-specific rubrics within each of these areas: Planning, Instruction, Assessment, Reflection, and Academic Language. The analytic scoring scheme is further shaped by a set of guiding questions in each area, as the following example shows (for elementary ELA):

## *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 (Pechione and Chung, 2006). This is the first study to examine PACT results in relation to student achievement.

## Literature Review

### Performance Assessment and Estimates of Teachers' Value-Added Achievement Gains

There are a number of areas of scholarship on teaching that are relevant to this study. First, rather than focusing on only one aspect of what defines a good teacher, it is better to consider multiple indicators of the quality of teachers, including measures of teacher quality, teaching quality, and teaching effectiveness (Darling-Hammond, 2007). In particular, this study seeks to expand on the literature on the relationship between these different ways of measuring what makes for a good teacher. Second, the proposed study explores the links between teacher quality and effectiveness and components of teacher education programs using large scale databases that incorporate many teacher education programs and many schools. Third, the study is the first to examine the relationship between teaching quality measured prior to the end of program by a teacher performance assessment and teacher effectiveness as measured in the classroom. It builds on research using performance assessments to determine National Board of Professional Teaching Standards (NBPTS) certification, which have typically found that such scores predict whether a teacher is more effective in raising student achievement. (Cantrell, Fullerton, Kane, & Staiger, 2008; Cavaluzzo, 2004; Goldhaber & Anthony, 2007; Vandevoort, Amrein-Beardsley, & Berliner, 2004).

Different authors have suggested that there are multiple ways to define and measure the quality of a teacher, because teaching always takes place within a specific context of students, subject matter, and location (Darling-Hammond, 2007; Campbell, Kyriakides, Muijs, & Robinson, 2004). It is helpful to make distinctions between concepts of teacher quality, teaching quality, and teaching effectiveness (Darling-Hammond, 2007). Such a distinction emerges partly from the way that each of these is measured. Teacher quality is the broadest of the three concepts, and refers to “the bundle of personal traits, skills, and understandings an individual brings to teaching, including dispositions to behave in certain ways” (Darling-Hammond, 2007). Teacher quality is composed of a set of qualities that would typically be relevant in a variety of contexts. For purposes of credentialing, for example, which certifies a teacher to teach in a variety of contexts, a teacher ought to be assessed on a breadth of knowledge and skills.

Teaching quality refers to the quality of a teacher's work in a particular context, with particular students. As such, it is narrower than teacher quality, involving contextualized practice to address specific conditions. More specifically, Darling-Hammond (2007) defines it in this way:

Teaching quality has to do with strong instruction that enables a wide range of students to learn. Such instruction meets the demands of the discipline, the goals of instruction, and the needs of students in a particular context. (p.5)

Teachers may be more adequate to teach well in certain subject areas and with certain students than in other contexts. For example, a teacher may be an expert at teaching young readers to read but not be so effective with helping older readers interpret Shakespeare. Teaching quality is best assessed by collecting data on practice with actual students. Furthermore, it is dependent to some extent on the conditions of teaching such as “curriculum materials, necessary supplies and equipment, reasonable class sizes, and the opportunity to plan with other teachers to create both appropriate lessons and a coherent curriculum across grades and subject areas” (p. 5) (Darling-Hammond, 2007). Performance assessments are measures of teaching quality within more or less controlled conditions, and are also indicators of the broader notion of teaching quality.

Teacher effectiveness is defined not by a teacher’s actions or skills but by the results for students. In some ways this is the most important criterion for defining a good teacher but it is in many ways the most problematic. First, teacher effectiveness is not entirely under the control of the teacher because students are also agents whose participation plays a key part of the learning process. Seemingly identical students may not choose to engage in the learning process to the same degree, regardless of what the teacher does. In addition, students have different supports for their learning at home and in the community, and bring different learning characteristics with them to the classroom. Finally, instructional conditions that matter for student learning (e.g. class sizes, curriculum materials, the availability of specialist supports, and so on) can vary both within and across schools.

Empirically, the primary challenge in assessing teacher effectiveness is the problem of isolating the contribution of the teacher to the outcomes by factoring out other influences that might affect a student’s outcomes. Value-added modeling using student achievement on standardized tests has gained in popularity because it shows some promise for statistically modeling a large number of student-level and school level variables. However, technical challenges and the reliance only on standardized tests make this a limited, if still useful, way to measure effectiveness.

These different ways to conceptualize and measure teacher effectiveness need to be kept in mind because while they are all related, they each can be measured in different ways and each is important in its own way. While a broad trait such as teacher quality is important for some contexts, it may not always be a good predictor of effectiveness in the classroom. Teaching effectiveness, a narrower concept, may provide additional information about whether, and in what contexts, a teacher can demonstrate good practices. Teacher effectiveness is the most important end goal for policy, but it is not always the easiest factor to measure and existing measures, such as students’ achievement gains, are also difficult to attribute with confidence to a single teacher, and are prone to error, especially at the individual level where samples of students are quite small.

Nonetheless, value-added measures of student learning can be useful for evaluating the effectiveness of larger groups of teachers, and connecting them to context-specific mea-

asures of teaching quality — such as those provided by performance assessments — may shed light on whether specific measures of what teachers do are related to measures of what students learn. That examination is what this study seeks to do, following on the heels of earlier studies of more mature performance assessments, such as the National Board for Professional Teaching standards portfolio and the Connecticut BEST portfolio.

### Studies of Teacher Performance Assessments as Measures of Teacher Effectiveness

A number of studies have explored whether National Board Certification and, more specifically, scores on the performance tasks used to determine such certification, is associated with teacher effectiveness as measured by student value-added achievement gains. With two exceptions finding mixed results, studies have found that higher scores on the NBPTS assessment are associated with raising student achievement. Overall, a number of studies have found evidence that performance assessments such as the National Board can predict teacher effectiveness, especially when the scaled scores on the test are used as predictors rather than just the determination of pass versus fail.

Goldhaber and Anthony (2007), studying student-level administrative data in North Carolina for the school years 1996-97 and 1997-98, used value-added models to see if National board certified teachers were more effective in increasing student achievement than other teachers. Their models included vectors of variables for individual students, teacher characteristics, school characteristics, district characteristics, and community characteristics. Modeling student-level and teacher level characteristics (without school level fixed effects) they found:

The magnitudes of the future NBCT coefficients suggest that student gains produced by NBCTs exceed those of noncertified applicants by about 4% of a standard deviation in reading and 5% of a standard deviation in math (based on a standard deviation of 9.94 on the end-of-year reading tests and 12.34 on the end-of-year math tests). (p. 141).

Thus, they found that NB certification, which depends on the results of a performance assessment-type portfolio of work, was associated with higher productivity of student achievement results. They wanted to explore whether this information provided by the NBC status was distinct from other information about teacher quality, namely the teacher's score on the state licensure exam. They were surprised to find that the coefficient for NB certification did not diminish much in its predictive power for teacher effectiveness when licensure score was added to the model, even though the licensure score was also a significant predictor of effectiveness. Based on this, they infer that, "NBPTS certification does in fact convey information about teacher quality above and beyond what can be learned from performance on teacher licensure tests alone" (p. 141). These results persisted even when they included school-level fixed effects.

Other studies have similarly concluded that National Board Certification is associated with higher effectiveness for raising student achievement. Vandevort, Amrein-Beardsley, and Berliner (2004) studying NBC teachers in Arizona, ran several analyses across three years and found that in most cases, NBC teachers showed higher student gains. Their model used prior year's scores as a covariate and did not include demographic factors, so it was based on a simpler statistical model than the other value-added models used to study National Board Certification and student achievement. Summarizing their results, they say:

In 75% (36/48) of the total cases, [effect sizes] were larger for students of NBCTs than for students of non-NBCTs. For the four years studied the effect sizes on the SAT-9, averaged across curriculum areas, were .203, .135, .037 and .112, yielding an overall average ES of about .122, indicating over one months gain per year on this standardized achievement test. (p. 34)

The statistical model used in this study, by not accounting for student level factors and fixed effects at the grade and school level, was less capable of addressing likely sources of bias, so these results are less definitive than some of those found in other studies. However, they nevertheless show similar trends.

Cantrell, Fullerton, Kane, and Staiger (2008) studied the relationship between teachers who had completed the National Board for Professional Teaching Standards (NBPTS) certification and student value-added achievement gains in math and English. Utilizing the scaled score for the NBPTS assessment and each of its ten sub-scores, they assessed which traits were related to value-added gains. In a school fixed effects model, the researchers found that the NBPTS overall scaled score was a significant predictor of teachers' value-added rating. While finding that successful NBPTS applicants differed significantly from unsuccessful NBPTS applicants, they did not find a difference between those who had applied for certification status and matched teachers (those randomly assigned to classrooms).

Whereas other studies have used elementary students, Cavalluzzo (2004) explored the differences in value-added for NBC teachers using grade 9 and grade 10 math results. After running various models that took into account teacher and student characteristics, including measures of student motivation, she found an advantage for the students of board certified teachers of about 0.07 standard deviations (p. 27).

A couple of studies have questioned the link between NBPTS assessment scores and teacher effectiveness, but in both cases the substantive findings differ little from those described above. Sanders, Ashton, and Wright (2005) compared value-added for students in English and math, grades 4-8, using data from North Carolina in the years 1999-00 through 2002-03, running several models for computing value-added, includ-

ing student demographic characteristics and teacher experience as well as National Board Certification. In some models they treated certification as a fixed effect and others, which they argued were superior, they treated certification as a random coefficient. They compared NBC teachers with all teachers, with those who failed the exam, and with those who planned to apply for NB certification in the future. Summarizing the fixed effects results, they write:

The sizes of the effects were generally less than one-half of a scale score unit and translated to standardized effect sizes that averaged 0.09 and 0.04 for math and reading, respectively, in Model 1, and 0.06 and 0.02 in Model 3. These effect sizes are roughly consistent with those reported by Cavalluzzo (2004) and by Goldhaber and Anthony (2004). (p. 6)

While results generally favored NBC teachers and were often significant in the fixed effects model, they were seldom significant in the random coefficient model.

Studying a comprehensive database of teachers in Florida, Harris and Sass (2007) found that NBC teachers outperformed others in teacher value-added estimates in some grades and subject tests, but not in others. By comparing NBC teachers with teachers who were not assessed for obtaining NBC status, this study does not provide as clear a focus on the ability of the NBC assessment itself to distinguish between more and less effective teachers as the other studies described above.

Finally, one study assessed the relationship between a study of beginning teacher performance assessment and teacher value-added in the state of Connecticut. Wilson, Hallam, Pecheone, & Moss (2006) used hierarchical linear modeling to study the relationship between performance on the Connecticut State Department of Education's Beginning Educator Support and Training (BEST) assessment system. Studying a sample of 110 3rd, 4th, 5th, and 6th grade teachers, and controlling for teacher background characteristics and performance on the paper-and-pencil Praxis II assessment, as well as student pretest and demographic characteristics, they found that performance on the portfolio was a statistically and substantively significant predictor of student achievement on the Degrees of Reading Power assessment. A one unit difference in BEST performance was associated with about 40% of a year's growth in reading. Thus, this study found very substantial predictive power in a performance assessment of beginning teachers.

Thus, performance assessments can provide predictive information about teacher effectiveness when used for veteran teachers taking the National Board for Professional Teachers exam. Such predictions can be statistically significant even when other teacher quality indicators are included in the model. This may indicate that performance-based information about teaching quality provides distinct information, or it may be that the information provided by performance assessments is similar to that provided by value-added assessments but, since each is still imperfect, more information is better. Such a pattern may generalize to preservice teachers such that performance assessments predict

future effectiveness. However, it is also possible that there might be a more restricted range of scores, which would make relationships more difficult to discern when following teachers into their teaching positions, especially if scores are used to determine whether teachers can graduate with a preliminary credential and can go on to teach. Furthermore, it might be that the relationship between teaching quality and teacher effectiveness differs for beginning teachers as compared with veteran teachers.

## Methods

### Sample

This study made use of administrative database of California teachers who were assessed on the Performance Assessment for California Teachers (PACT) and a database for a large urban school district that links teachers and students (but for only a subset of the cases). Because we are still addressing the challenge of linking pre-service teachers with district databases, the following analyses are preliminary, and we expect to increase the sample of teachers and districts in subsequent analyses. This initial exploratory study was conducted largely from candidates from a district intern program as well as one candidate from a traditional teacher education program in the same district.

We established links between the PACT scores and district teachers in two ways: district interns were designated with PACT IDs that matched their employee IDs, year of PACT assessment, and year beginning in the district. Using this method 34 teachers who took the elementary English language arts PACT assessment were reliably linked to their PACT scores. PACT participants from a district intern program had district employee IDs listed as their PACT IDs, which allowed us to be reliably able to link intern data with district data based on their IDs, and then to link these teachers to their students. For traditional program participants, we linked teachers based on name, year of PACT assessment, and first year with the district. We excluded teachers with fewer than five linked students because of the unreliability of estimating teacher effects with so few students, and teachers who were not linked to students in the data set. We also included only teachers who took the Elementary English PACT assessment, and teachers who we could link by name and start date on both the PACT database and the university database. Ultimately, the analysis was based on 259 students of 14 teachers in grades 3-6.

### Value-Added Estimates

Estimates of teachers' value-added were calculated for all multiple-subjects teachers in the district over the years 2003-04 to 2007-08. Four value-added models were used in order to assess whether the results were robust across different modeling choices. All models used ordinary least squares (OLS) regression to predict student achievement in ELA from ELA achievement in the prior year for grades 3-8 in school years 2003-04 through 2007-08, with all achievement scores standardized by grade/test level. The simplest model included only achievement regressed on prior achievement. A second

model added student demographic variables (race/ethnicity, sex, special education status, and whether the student had been retained in grade<sup>1</sup>) to this basic model. The other two models added school-by-year fixed effects to these first two as a way to control for school effects on learning. These models were then used to produce a residual score for each student, essentially comparing his or her actual score to the predicted score. Teacher value-added was computed as the mean of these residual scores for each teacher's students across years.

### *Analytic Approach*

In order to examine the relationship between teacher performance on PACT and the value-added gains in student achievement associated with each teacher, we first aggregated the residual scores for students of each teacher to obtain an estimate of teacher value-added and then correlated these value-added ratings with PACT scores. Value-added analyses use statistical models used to compare outcomes of a teacher's students with the outcomes for similar students in the sample, in this case, a large urban school district. Such models allow for estimating a teacher's effects on student learning relative to other teacher's in the sample. It is generally agreed that value-added estimates provide more useful information about a teacher's effects on student learning than data on student outcomes alone because they address factors such as prior achievement and student demographic characteristics that can have a large influence on student achievement at a given time. However, even the best models fall short as causal estimates, for a variety of reasons (Rubin, Stuart, & Zanutto, 2004), and so we explored a number of alternative models as a check of whether the findings were robust against model choice.

### *Results*

Table 1 (following) summarizes the range of scores PACT components, subtests, and overall scores. Among the subscales, *planning* and *instruction* items showed the highest scores and the least variability, whereas *assessment* and *reflection* items had the lowest mean scores and the greatest variability.

---

1. We would also have liked to include EL status, achievement on the California English Language Development Test, and parent education, but we did not have complete information about these variables.

Table 1: Description of Candidate Performance on PACT Components (N=14)

| PACT Component           | Mean         | Std. Dev.   | Min       | Max       |
|--------------------------|--------------|-------------|-----------|-----------|
| <b>Planning</b>          | <b>10.93</b> | <b>1.38</b> | <b>8</b>  | <b>12</b> |
| p1                       | 3.64         | 0.50        | 3         | 4         |
| p2                       | 3.71         | 0.61        | 2         | 4         |
| p3                       | 3.57         | 0.51        | 3         | 4         |
| <b>Instruction</b>       | <b>7.03</b>  | <b>1.18</b> | <b>4</b>  | <b>8</b>  |
| i1                       | 3.64         | 0.63        | 2         | 4         |
| i2                       | 3.39         | 0.62        | 2         | 4         |
| <b>Assessment</b>        | <b>5.50</b>  | <b>1.51</b> | <b>4</b>  | <b>8</b>  |
| a1                       | 2.93         | 0.73        | 2         | 4         |
| a2                       | 2.57         | 0.85        | 2         | 4         |
| a3 (optional)            |              |             |           |           |
| <b>Reflection</b>        | <b>6.00</b>  | <b>1.84</b> | <b>3</b>  | <b>8</b>  |
| r1                       | 2.86         | 1.10        | 1         | 4         |
| r2                       | 3.14         | 0.86        | 2         | 4         |
| <b>Academic Language</b> | <b>6.71</b>  | <b>1.14</b> | <b>4</b>  | <b>8</b>  |
| al1                      | 3.07         | 0.73        | 2         | 4         |
| al2                      | 3.64         | 0.63        | 2         | 4         |
| <b>Total PACT Score</b>  | <b>36.17</b> | <b>5.46</b> | <b>24</b> | <b>44</b> |

Table 2 summarizes the value-added estimates for teachers, which reflect the residual, or difference, between their average student performance (expressed as an aggregated standardized z score) and the average performance of similar students in the school district. Somewhat surprisingly, the students of these beginning teachers outperformed similar students in the school district, since the average student residual is positive and centered very close to 0. They showed a considerable range of scores regardless of the statistical model.

Table 2: Early Teacher ELA Value-Added Estimates

| Statistical Model                       | Mean | Std. Dev. | Min   | Max  |
|---|------|-----------|-------|------|
| Pretest Only                            | 0.07 | 0.28      | -0.37 | 0.68 |
| Pretest Plus Demographics               | 0.08 | 0.30      | -0.33 | 0.67 |
| Pretest Only, School Fixed              | 0.10 | 0.28      | -0.26 | 0.64 |
| Pretest Plus Demographics, School Fixed | 0.07 | 0.31      | -0.31 | 0.62 |

Table 3 (page 13) summarizes the correlations between value-added for each model with scores on PACT items, subscales, and total score. In general, the choice of value-added model made at most a modest difference in the sizes of the correlations, so the results do not appear, in most cases, to be model-dependent. Overall, PACT scores were correlated about .5 with teacher value-added (with correlations ranging from .46 to .53, depending on the model), indicating a moderately strong relationship. The magnitude of the relationship between PACT total score and student value-added achievement gains was quite large. For example, in the model using pretest only, a one standard deviation difference in PACT score was associated with a 0.15 difference in teacher value-added (measured in standard score units). This is about twice the size of the coefficients typically estimated for board certified vs. matched teachers.

Another way of evaluating this effect size is that, for each point a teacher scored on PACT (scored on a 44-point scale), her students averaged a gain of about one percentile per year compared with similar students. For each additional point a teacher scored on PACT, her students averaged a gain of one percentile point per year on the California Standards Tests as compared with similar students. Thus, 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). Given the small sample size, these effects were slightly above the traditional cut-off for statistical significance of  $p < 0.05$  level, with  $p$  values ranging from 0.053 to 0.087.

Several subscales were strongly correlated with teachers' value-added estimates. Assessment was the most consistent predictor of value-added, showing statistical significance across all value-added models, with the subscale as a whole correlated about .6, and the first guiding question score correlated about .7 with value-added scores. One of the indicators of teachers' ability to develop academic language (a combination of ability to develop English language proficiency for limited English proficient students and to develop academic language within the content area) was also strongly related with value-added, with 3 of the 4 correlations reaching .8. In this case, not surprisingly, the strongest relationships were found using models that controlled for student demographics and school effects (which would take into account school demographics, among other things). Instruction had a small correlation with value-added of around .3 for the subscale. Planning, which had the highest and least variable scores, had little predictive relationship with value-added.

Table 3: Correlations Between PACT Performance and Teacher Value-Added

| PACT Component           | Statistical Model for Estimating Value-Added |  |                                     |   |
|--------------------------|--|--|-------------------------------------|---|
|                          | (1) Prior Test Only, no School Fixed         | (2) Demographics and Prior Test, no School Fixed | (3) Prior Test Only, W/School Fixed | (4) Demographics and Prior Test, W/School Fixed |
| <b>Planning</b>          | <b>0.06</b>                                  | <b>0.01</b>                                      | <b>0.03</b>                         | <b>0.01</b>                                     |
| p1                       | 0.06   | 0.00   | 0.07                                | 0.04  |
| p2                       | 0.00   | -0.05  | -0.03                               | -0.05   |
| p3                       | 0.11   | 0.08   | 0.05                                | 0.05  |
| <b>Instruction</b>       | <b>0.34</b>                                  | <b>0.29</b>                                      | <b>0.30</b>                         | <b>0.29</b>                                     |
| i1                       | 0.29   | 0.24   | 0.20                                | 0.19  |
| i2                       | 0.34   | 0.31   | 0.37                                | 0.35  |
| <b>Assessment</b>        | <b>0.60*</b>                                 | <b>0.58*</b>                                     | <b>0.68**</b>                       | <b>0.66*</b>                                    |
| a1                       | 0.73**                                       | 0.69**   | 0.75**                              | 0.72**  |
| a2                       | 0.45   | 0.44   | 0.56*                               | 0.55*   |
| a3 (optional)            | .  | .  | .                                   | .   |
| <b>Reflection</b>        | <b>0.50+</b>                                 | <b>0.45</b>                                      | <b>0.30</b>                         | <b>0.29</b>                                     |
| r1                       | 0.46+  | 0.40   | 0.28                                | 0.25  |
| r2                       | 0.49+  | 0.45   | 0.29                                | 0.30  |
| <b>Academic Language</b> | <b>0.49+</b>                                 | <b>0.47+</b>                                     | <b>0.62*</b>                        | <b>0.61*</b>                                    |
| al1                      | 0.52+  | 0.53+  | 0.79**                              | 0.77**  |
| al2                      | 0.29   | 0.24   | 0.20                                | 0.19  |
| Total PACT Score         | 0.53+  | 0.48+  | 0.49+                               | 0.47+   |

+ =  $p < .10$  \* =  $p < .05$ , \*\* =  $p < .01$

In summary, these preliminary findings suggest that teacher performance assessment on the PACT may provide useful information about which teachers are likely to most contribute to students' value-added achievement. If these findings are confirmed in the larger scale studies that are underway, they would suggest the validity of PACT as a measure of teacher quality, and as a useful tool for evaluation of teacher candidates and as a way to provide feedback to teacher education institutions.

## References

- Amrein-Beardsley, A. (2008). Methodological concerns about the education value-added assessment system. *Educational Researcher*, 37(2), 65-75.
- 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.
- 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., 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*. Mahwah, 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.
- Harris, D., and Sass, T.R. (2007). The effects of NBPTS-certified teachers on student achievement. Unpublished manuscript. Downloaded 10 October, 2010 from [http://www.nbpts.org/userfiles/file/harris\\_sass\\_final\\_2007.pdf](http://www.nbpts.org/userfiles/file/harris_sass_final_2007.pdf)
- 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.
- 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.
- 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](http://www.nbpts.org/UserFiles/File/SAS_final_NBPTS_report_D_-_Sanders.pdf)
- Vandevoort, L. G., Amrein-Beardsley, A. & Berliner, D. C. (2004). National board certified teachers and their students' achievement. *Education Policy Analysis Archives*, 12(46), 1-117.
- Wilson, M., Hallam, P. J., Pecheone, R., and Moss, P. (2006). Using Student Achievement Test scores as evidence of external validity for indicators of teacher quality: Connecticut's Beginning Educator Support and Training program. Unpublished manuscript.

## Appendix A: Summary of Value-Added Models

| Predictor <sup>1</sup>     | Model 1            | Model 2 <sup>2</sup> | Model 3             | Model 4 <sup>4</sup> |
|----------------------------|--------------------|----------------------|---------------------|----------------------|
| Prior ELA (standardized)   | 0.857**<br>(0.001) | 0.799**<br>(0.001)   | 0.800**<br>(0.004)  | 0.759**<br>(0.004)   |
| Ethnicity Asian            |                    | 0.169**<br>(0.004)   |                     | 0.150**<br>(0.004)   |
| Ethnicity African-American |                    | -0.002<br>(0.003)    |                     | -0.017<br>(0.010)    |
| Ethnicity Filipino         |                    | 0.127**<br>(0.004)   |                     | 0.136**<br>(0.014)   |
| Ethnicity Native American  |                    | 0.105**<br>(0.015)   |                     | 0.143*<br>(0.050)    |
| Ethnicity Pacific Islander |                    | 0.047**<br>(0.010)   |                     | 0.042<br>(0.030)     |
| Ethnicity White            |                    | 0.192**<br>(0.003)   |                     | 0.128**<br>(0.010)   |
| Ethnicity Other            |                    | 0.142**<br>(0.018)   |                     | 0.083<br>(0.060)     |
| Female                     |                    | 0.044**<br>(0.044)   |                     | 0.042<br>(0.006)     |
| Special Education          |                    | -0.185**<br>(0.003)  |                     | -0.219**<br>(0.010)  |
| Retained                   |                    | 0.165**<br>(0.016)   |                     | 0.233**<br>(0.051)   |
| Constant                   | 0.067**<br>(0.001) | -0.004+<br>(0.002)   | -0.187**<br>(0.068) | -0.215**<br>(0.067)  |
| School Fixed?              | No                 | No                   | Yes                 | Yes                  |
| Mean Squared Error         | 0.53               | 0.52                 | 0.53                | 0.53                 |
| Adjusted R-Squared         | 0.72               | 0.73                 | 0.72                | 0.72                 |

\*=p<.05, \*\*=p<.01

1. Dependent variable is ELA achievement on the California Standards Test (CST), standardized by grade/test.

2. These models included all of the same students of PACT teachers as Models 1 and 2, but, because of missing data for many district students, the models were based on approximately 30,000 cases whereas the other models included approximately 280,000 cases.

**Stanford Center for Assessment,  
Learning, and Equity**

Barnum Center, 505 Lasuen Mall

Stanford, California 94305

Phone: 650.725.0703

<http://scale.stanford.edu>