Assessing Predictive Validity of Teacher Performance Assessment (TPA) for Teaching Credential Candidates: A Case Study

Tseh-sien Kelly Vaughn
Kim Tolley
Stephanie Demaree

Notre Dame de Namur University
AERA 2014

Abstract

This quantitative study evaluates the predictive validity of Teacher Performance Assessment (TPA) at Notre Dame de Namur University, located in Belmont, California. We examined the relationship between 80 teaching credential candidates’ scores on the Performance Assessment for California Teachers (PACT) in the 2012-13 academic year and three other measures of candidate effectiveness: California Basic Educational Skills Test (CBEST) scores; Grade Point Averages (GPAs); and university field-based assessment of teacher performance known as the Master Teacher /Supervisor Evaluation (MTSE). Data that resulted from conducting a Pearson’s test of correlation reveals PACT has a moderate predictive validity when aggregate scores are considered, but no predictive validity between individual students’ mean PACT scores and their mean GPA, CBEST, or MTSE scores.
Objectives and Purpose

This study investigates the relationship between teaching credential candidates’ scores on the Performance Assessment for California Teachers (PACT) and three other measures of candidate effectiveness: 1) California Basic Educational Skills Test (CBEST) scores; 2) Grade Point Averages (GPAs); 3) Notre Dame de Namur University’s (NDNUs) field-based assessment of teacher performance: the Master Teacher /Supervisor Evaluation (MTSE).

A number of scholars have recently called for additional research into the predictive validity of teacher performance assessment (Petroff & Whittaker, 2011; Pecheone, 2012; Riggs, Verdi & Arlin, 2009; Sandholtz, 2012). Our study aims to build upon previous inquiry and research by investigating the relationship between NDNU credential candidates’ PACT scores and other academic and field-based assessment results.

Perspective or Theoretical Framework

It is important to establish the predictive validity of Teacher Performance Assessments (TPAs), as their implementation has policy implications. A recent federal initiative has addressed teacher quality by connecting the readiness of teacher candidates with the results of their TPAs (U.S. Department of Education, 2011, p. 10). This is congruent with recent national and statewide organizations that have taken similar steps (Bryk, Harding, & Greenberg, 2012). The trend towards assessing the readiness of pre-service teachers has led to the support of TPAs, most notably the development and implementation of edTPA by SCALE, or the Stanford Center for Assessment, Learning and Equity (2012). The current state and national reform initiatives reflect the majority of research using TPAs as summative, which are formal measurements of pre-service teacher candidate readiness. In California, candidates that receive failing scores on
the TPA cannot earn a teaching credential. The TPA thus serves as a high-stakes gate-keeping mechanism.

To address predictive validity, some scholars have outlined particular measures to assess TPAs. In 2011, Susan Petroff and Andrea Whittaker emphasized the importance of evidence to support TPA use for accreditation and licensure decision-making. Their proposed research agenda included analysis of the reliability between performance on the TPA and other measures, such as state teacher certification test scores. In 2012, Raymond Pecheone of SCALE proposed a research agenda that included clarification of the relationship of TPE to GPA, content knowledge examinations, value-added metrics and other measures.

There are also recent studies that address TPA predictive validity by examining the relationship of TPA scores and another measure in the candidate’s teacher education program. In 2009 study of the validity and reliability of TPAs at California State University, San Bernardino, Matthew L. Riggs, Michael P. Verdi, and Patricia K. Arlin recommended that researchers move away from comparing TPA assessment to more academic measures such as GPA and test scores; and begin comparing TPA assessment to more comparable measures of teaching performance such as master teacher and/or supervisor evaluations. A more recent 2012 study by Judith H. Sandholtz and Lauren M. Shea compared supervisors’ predicted PACT scores to actual scores in order to assess the validity of this TPA to the formative evaluations done by supervisors. In a follow up study, Sandholtz (2012) further examined candidates with the greatest differences between the supervisor’s predictions and actual PACT scores. This body of research informs our study.
Methods, Techniques, or Modes or Inquiry

This study builds upon previous research by comparing Notre Dame de Namur University (NDNU) candidates’ PACT results with other assessment constructs, including GPA, CBEST, and NDNU’s MTSE. Our research addresses the following questions: (1) What is the correlation and predictive validity of PACT results with other forms of assessment used in NDNU’s Multiple and Single Subject Credential Programs? (2) How can the relationship of PACT with other assessment measures inform our policy and practice as teacher educators?

The population for this study included 80 students representing all the candidates that had completed the PACT in the 2012-13 academic year. Data analysis involved quantitative research methods. Our analysis yielded descriptive statistics, correlation coefficients, and the probability of the results.

Data Sources

Data were collected from August 2012 through May 2013. Four data sources were used, as outlined in the following:

1. PACT: This TPA was designed to measure the candidate's knowledge, skills and ability related to California's Teaching Performance Expectations (TPEs) (California Commission on Teacher Credentialing, 2013). It is currently one of three TPAs approved by California Commission on Teacher Credentialing (CCTC) as requirement for all multiple (elementary) and single (secondary) subject teaching credentials to pass a TPA for teacher licensure (CCTC, 2006; Merino & Pecheone, 2013). PACT encompasses five tasks that a candidate needs to complete and submit altogether as a teaching event. A trained and calibrated scorer then holistically scores the teaching event using 12 rubrics. Scorers mark each rubric by declaring a score from one to four and providing evidence to justify the score. A level one score represents a candidate that is
not ready to teach but shows some understanding of the criteria; a level two score represents a candidate that is ready to teach with support; a level three score represents a candidate who is ready to teach independently; and a level four score represents candidate who is an exceptional teacher. A candidate who receives three or more scores of one fails the PACT (Pecheone & Chung Wei, 2007, p. 12).

2. GPA: Notre Dame de Namur University determines the graduate student's progress by assessing the academic and professional behavior of the student by means of letter grades. The grade point average is computed on a four-point system, allowing for plus and minus grades. In the Credential Programs, students must maintain a cumulative GPA of 3.0 or higher.

3. CBEST: This assesses a candidate’s basic content understanding in reading, writing, and mathematics. As such, the CBEST does not measure a candidate’s ability to teach this basic content but rather documents the requirements relating to credentialing and teacher employment (Pearson Education, 2012).

4. MTSE: At NDNU, this is a field-based assessment of the candidate’s work in his/her student teaching assignment. The evaluation instrument assesses effectiveness according to 13 Teacher Performance Expectations (TPEs) identified by the California Commission on Teacher Credentialing. Master Teachers are highly experienced, having taught for a minimum of five years. During the semester, both the Master Teacher and the University Supervisor provide the teacher candidate with regular feedback concerning his/her progress. The University Supervisor is required to complete a Teacher Observation report at each visit and present one copy to the student teacher/intern. The Master Teacher may also keep written records of observations and conferences. Near the end of the semester, the Master Teacher, student teacher/intern, and the University Supervisor provide a cumulative assessment of the student teacher/intern’s
effectiveness by scoring each TPE on the MTSE. The MTSE also allows space for comments from the Supervisor and the Master Teacher. For this study we used candidates’ final evaluations submitted at the end of their second semester student teaching (NDNU, 2013).

Results

The predictive validity of the PACT was investigated by conducting a Pearson’s test of correlation between student cumulative mean performance on PACT, GPA, CBEST scores, and MTSE scores.

Because the development of the PACT portfolio involves a substantial amount of written work, we expected that some degree of correlation might exist between performance on the PACT and other academic measures such as GPA and the CBEST. We found no correlation between performance on the PACT and the CBEST. We also analyzed the CBEST sub-category scores, hypothesizing that there might be a moderate correlation between students’ CBEST writing scores and PACT scores. The resulting analysis did not confirm this expectation. However, we did find a very moderate correlation between students’ cumulative grade point averages and their cumulative PACT scores (see correlations in Table 1).

Table 1: Correlations between PACT, GPA, CBEST, and MTSE Cumulative Mean Scores. N = 80

|                | PACT  
|----------------|-------
| GPA            | 0.308*|
| CBEST Cumulative | -0.279|
| CBEST Reading Scaled | -0.22 |
| CBEST Math Scaled    | -0.205|
| CBEST Writing Scaled | -0.169|
| MTSE            | -0.054|

* p < .01
Because NDNU’s MTSE provides a more comparable measure of teaching performance to the PACT than do GPA or CBEST scores, we expected to find predictive validity between the two assessment instruments. A Pearson’s test of correlation was conducted to compare the candidates’ mean composite scores on the MTSE to their mean composite PACT scores.

We found no correlation -- and thus no predictive validity -- between individual students’ mean PACT scores and their mean MTSE scores. Both assessment instruments relate to TPEs and are based on actual performance in the classroom, and so this finding was somewhat unexpected.

In addition to comparing overall cumulative mean scores, we also compared sub-category scores. The PACT weights its assessment categories differently than MTSE. For example, the PACT task category *Planning-Establishing a Balanced Instructional Focus* is related to TPEs 1, 4, and 7, and the PACT task category *Instruction-Engaging Students in Learning* is based on TPEs 1, 5, and 11. To better match the weightings in subject areas, we recalculated the MTSE scores to match the PACT weightings in these two categories. However once again, we found no correlation between PACT scores and MTSE scores, (see Table 2).

Table 2: Correlations between PACT and Re-weighted MTSE Category Task Scores

<table>
<thead>
<tr>
<th>PACT Teaching Event Task</th>
<th>Related TPEs</th>
<th>PACT Mean Scores</th>
<th>MTSE Mean Scores, Re-weighted</th>
<th>Pearson’s r</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planning – Establishing a Balanced Instructional Focus</td>
<td>1, 4, 9</td>
<td>2.81</td>
<td>2.89</td>
<td>0.081</td>
</tr>
<tr>
<td>Instruction – Engaging Students in Learning</td>
<td>1, 5, 11</td>
<td>2.58</td>
<td>2.86</td>
<td>0.086</td>
</tr>
</tbody>
</table>

Data suggests that there is no correlation between the individual candidate scores yielded by PACT and the individual scores yielded by the MTSE. In other words, a candidate’s overall
score on the PACT does not predict that student’s overall score on the TPE MTSE, and vice versa.

On the other hand, we did find a moderate correlation ($r = 0.50; p < .001$) between the overall mean scores of the student population within each criterion considered within the PACT and MTSE assessments. In other words, aggregate mean high or low scores in specific PACT criteria will predict a similar pattern of mean high or low scores in the same criteria assessed by the MTSE. Thus, a low mean score in the area of “Reflecting on Learning” in the PACT assessment will predict a comparable low mean score in the area of “Reflection and Feedback” on the MTSE.

**Significance of the Study**

We found that PACT has a moderate predictive validity only when aggregate scores are considered. We conclude that aggregate PACT scores can be useful tools for summative evaluation of the credential programs. They generally accord reasonably well with the MTSE evaluations, identifying overall areas of candidate strength and weakness that can be addressed through program improvement. PACT scores can be used to make programmatic changes at NDNU. In fact we have been using the scores in this way for some time. However, the lack of predictive validity when it comes to individual student performance on the PACT and MTSE raises some questions. It will be important in the future to explore the cause of the discrepancies between the PACT and MTSE scores of individual candidates and understand how these might be addressed.

As stated, we found that PACT has no predictive validity when it comes to individual student performance. Given the high-stakes nature of PACT, this finding is troubling. One possible explanation is that the PACT includes only a 10-20 minute video snapshot of actual
teaching, whereas both Supervisors and Master Teachers have many hours of direct observation of the teacher candidates. The Master Teachers have 10 hours of contact with single subject candidates and 20 hours with multiple subject candidates weekly to observe the TPEs. In contrast, the PACT is a portfolio in which candidates must provide evidence of meeting the TPEs. A follow-up study could be conducted to survey the Master Teachers, most of whom are not calibrated PACT scorers, to uncover the reasons for discrepancies between the MTSE and PACT scores.

Given these findings, it is clear that more work needs to be done to explore the relationship between TPAs like the PACT and other forms of field-based assessments like NDNU’s MTSE. This may help to establish the predictive validity of TPAs.
References


