

Toward a Culture of Evidence: Factors Affecting Survey Assessment of Teacher Preparation

**By Paul Beare, James Marshall,
Colleen Torgerson, Susan Tracz, & Robin Chiero**

Introduction

Developing a culture of evidence to assess and improve teacher preparation programs is a critical issue in American education. Teacher education has been struggling with the challenge of preparing and retaining sufficient numbers of high-quality teachers who can work effectively with students from all cultural and racial backgrounds, raising the achievement for all students (Wang, Spalding, Odell, Klecka, & Lin, 2010). Darling-Hammond (2002) found that teacher preparation

is a stronger correlate of student achievement than class size or school spending, accounting for 40% to 60% of the variance in achievement. Teachers who learn and practice sound pedagogical techniques can affect students' measured achievement (Blair, 2000). Although these studies indicate that teacher quality is the most important factor influencing student achievement (Whitehurst, 2002), even among those who believe the high quality preparation of teachers is critical, there are sharp contrasts concerning the best approach (Levine, 2006).

Many scholars suggest that a strong research base on how best to prepare teachers to meet the challenges

Paul Beare is a professor and dean and James Marshall, Colleen Torgerson, Susan Tracz, and Robin Chiero are professors, all with the Kremen School of Education and Human Development at California State University, Fresno.

Toward a Culture of Evidence

of today's classrooms is lacking (Wilson, Floden, & Ferrini-Mundy, 2001). Metzler and Blankenship (2008) discussed a "paucity of systematically collected evidence" in teacher preparation assessment despite it being central to the conduct and future of teacher education (p. 1098). Cochran-Smith (2003) posited that formal program assessment efforts are noticeably lacking in teacher education. This shortage of evidence results in a myriad of potential "solutions" regarding teacher preparation, but with few ways to evaluate their promise (Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2008).

Numerous reports and analyses have focused on this lack of a research base with most demanding better and more authentic assessment (Darling-Hammond, 2006). Concurrently there is a national demand for the reform of teacher education, particularly university-based preparation (Capraro, Capraro, & Helfeldt, 2010). Educational coursework has been found to have a critical point of diminishing returns and several studies have indicated that teachers with advanced subject matter degrees, rather than advanced education degrees, produce students who perform better in math and reading (Kaplan & Owings, 2002). A credential in education may be sufficient to produce student learning, but greater content knowledge has been found to affect learning as much as advanced education degrees (Greenwald & Hedges, 1996).

Eleven years ago Zeichner (1999) pointed out that education faculty must do the best job possible in preparing teachers for our schools or perhaps let someone else do the job. Many voices echo that sentiment, including Secretary of Education Arne Duncan (2010) who asserted that many of the nation's 1,450 schools, colleges, and departments of education are doing a mediocre job of preparing teachers for the realities of the 21st century classroom. This type of change requires quality assessment and a clear understanding of what the resulting data indicate.

The evidence-based education movement, which holds that decisions about practice and policy should be made on the basis of empirical evidence about outcomes, is now predominant (Moss, 2007), despite the defensiveness and recalcitrance of some faculties of education (Akmal & Miller, 2003). Many initiatives are intended to create new cultures of evidence or inquiry in institutions (Knapp, Copland, & Swinnerton, 2007) and/or to "re-culture" organizations so that using evidence and assessment data becomes central to the way decisions about local policy and practice are made (Louis, 2008). Cochran-Smith (2009) called for new cultures of evidence and inquiry in teacher education and stated that they have the potential to be transformative and revitalizing. She also pointed out that current discussions about creating cultures of evidence in teacher preparation often do not reflect the understanding of culture or its resistance to change. Gee (2007) stated that in assessment of teacher preparation there is a conspicuous absence of cultural nuance, including an absence of situated understandings of the role of human interpretation in constituting and using evidence. While many reports discuss the need to rely on evidence in making programmatic decisions, there is little discussion about how such a system would coincide with the local cultures of colleges and universities (Cochran-Smith, 2009).

Cochran-Smith and Zeichner (2005) discussed the difficulty of the research needed to improve teacher preparation programs and pointed out that it depends on several critical links which could connect teacher preparation programs with candidate knowledge, skills, and dispositions, and the candidates' actual practices in classrooms, eventually linking to pupil learning while in the graduates' charge. They stated,

. . . unraveling the complicated relationships between and among these variables and the contexts and conditions in which they occur is exceedingly complex, and of course this entire enterprise assumes in the first place that there is consensus about appropriate and valid outcome measures, an assumption that is arguable. (p. 3)

What Is Valid Evidence?

The Teachers for a New Era initiative proposed a conceptual framework for the development of an evidence portfolio to demonstrate and assess a teacher education program's success in preparing teachers (Cochran-Smith, 2009). While there are multiple components to such a portfolio, the first area of importance is a survey and tracking of graduates. Blanton, Sindelar, and Correa (2006) identified large-scale surveys, teacher checklists, and comparison to standards as three of the five ways in which beginning teacher preparation quality may be examined. There is general agreement that a teacher graduate's effect on student achievement is an important variable to examine a program's effectiveness, yet no proven methodology exists for accomplishing this.

The "value added" approach in particular is under attack (Baker et al., 2010). Teacher performance assessments using teacher work sample methodology developed by The Renaissance Group are useful for examining individual teacher's effects on achievement (Torgerson, Macy, Beare, & Tanner, 2009) but have not been used systematically for program evaluation. Despite these evidence-oriented initiatives, little has been done to evaluate the quality of the evidence being generated or develop systematic ways to use that evidence to improve teacher preparation (Ludlow et al., 2008).

Present Research

This study investigated the effect selected extrinsic variables have on survey data collected to determine the efficacy of, and improve, teacher preparation programs. While recognizing other aspects of program evaluation, isolating the effects of extrinsic variables on the survey results is an important step to determining whether the results can be accepted at face value or if they are influenced by outside factors over which programs have no control.

In working toward a culture of evidence concerning teacher preparation, all schools, departments, and colleges of education of the 23 California State University (CSU) campuses established common assessments as recommended by Cochran-Smith (2009) and Darling-Hammond (2006). In 1999, a survey of credentialed

Toward a Culture of Evidence

graduates at the end of their first year of professional teaching and the graduate's employment supervisor during that year of teaching was initiated by the CSU Education deans and the system Chancellor. The survey contains specific questions about the quality of preparation provided by the credential program. Each campus receives an annual report from the CSU Center for Teacher Quality (CTQ) with results from the survey concerning the previous year's graduates and supervisors. The report also includes a summary of all data since the inception of the surveys for comparison purposes, and parallel results for the 23 CSU campuses compiled system-wide. This unique service allows each campus to track the effects of program changes designed to improve performance.

As Ludlow et al. (2008) predicted, many campuses have struggled to develop systematic ways of using this rich body of evidence to improve teacher preparation. Teacher education faculty are well aware of the complex web of variables described by Cochran-Smith and Zeichner (2005). They often question the survey results, citing extrinsic factors to explain differences between the scores obtained by their own program versus those obtained elsewhere on campus or in the greater CSU (Beare, 2009). Beliefs concerning some factors are based on what "a priori" may seem important, (such as, the number of university credits required or the number of students in a program) and some are based on conventional wisdom concerning important K-12 school characteristics (e.g., socio-economic status, English Language ability of students, or school achievement level). This study examined two factors specific to the preparation programs, and four extrinsic factors specific to the K-12 schools in which surveyed graduates were teaching.

Survey Instrument

The *Systemwide Evaluation of Professional Teacher Preparation Programs* (SEPTPP) compiles evidence about the extent to which K-12 teachers who are recent graduates of credential programs on CSU campuses are prepared for their most important teaching responsibilities and the extent to which CSU professional coursework and fieldwork were professionally valuable and helpful to them during their initial year of K-12 teaching (CTQ, 2009). This is accomplished by asking both graduates and the graduates' employment supervisors to complete separate, but parallel, 110-item online surveys at the end of the graduate's first year of full time professional teaching employment.

The instrument includes common questions for all teachers and supervisors as well as credential-specific questions for particular groups. They are queried about the extent to which the teachers were prepared for important responsibilities that are commonly assigned to K-12 teachers. Teachers are also asked common questions about the extent to which major features of the preparation programs proved to be valuable and helpful during subsequent teaching. Finally, respondents reply to questions about the quality of the credential programs in relation to prominent standards for state and national accreditation. Teachers and supervisors are also asked

credential-specific questions about the extent to which the teachers were prepared for responsibilities that are most commonly associated with their specific assignment (CTQ, 2009). Supervisors are asked to evaluate new teachers as novices, and only if they have observed and had a conference with them during this first year.

Data Collection

Each CSU campus forwards to the CTQ a list of former students at that campus who, during a prior 12-month period, met the standards for state certification as K-12 teachers. School sites are identified for approximately 55% of the completers from these sources. After receiving an initial list, the CTQ and CSU campuses make a second effort to find the school locations of additional teachers by directly contacting approximately 1,000 school districts and 50 county offices of education. This effort yields site information for an additional 30 percent of recent CSU completers (CTQ, 2009).

Validity of the Evaluation

The validity of the evaluation derives from the alignment between the evaluation questions and (1) California standards for grades K-12 in all curriculum areas, (2) California Standards for Accreditation of Professional Teacher Preparation, (3) California Teaching Performance Expectations, (4) California Standards for the Teaching Profession, and (5) Standards adopted for institutional accreditation by the National Council for Accreditation of Teacher Education (CTQ, 2009). Individuals who had participated in drafting and implementing California's accreditation standards for universities and its performance expectations for teachers were responsible for the alignment of the evaluation questions (CTQ, 2006).

Reliability of the Evaluation

Uncertainty about evaluation findings comes from two principal sources, the number of evaluation participants and the extent of their concurrence with each other. The evaluation findings become increasingly certain to the extent that the questions are answered by increasing numbers of program completers and their employment supervisors. Each year the data set yields the percent of respondents who gave specified answers to each item and includes reliability estimates in the form of confidence intervals based on the number of respondents and the concurrence or homogeneity of responses. In 2003, the CSU Deans of Education grouped together substantively related evaluation questions into "composites." For example, the survey includes several questions about preparing teachers for diversity in education. The deans grouped these questions together in a composite called *Preparing for Equity and Diversity in Education*. These groupings facilitate the analysis and interpretation of large amounts of complex data and the composite scores are substantially more reliable than are the participants' responses to individual survey questions and are sufficiently valid and reliable to serve as the basis for academic

Toward a Culture of Evidence

and professional decisions about teacher preparation (CTQ, 2006). The reliability for the composite scores for the system and the individual campuses generally range from 0 to 2 percentage points at the 90% confidence level.

Research Question

Do specific extrinsic variables significantly influence results of surveys evaluating CSU teacher preparation programs? This concern stated by Ludlow et al. (2008) has been omnipresent in discussions of the CSU survey data by faculty and deans. Both groups question if extrinsic variables are responsible for survey outcomes instead of or in addition to the actual preparation program strengths or weaknesses in specific assessment areas (Beare, 2009). To address this question, the present research examined the effect of specific extrinsic variables on the preparation program ratings by supervisors and graduates across the 23 campus CSU system.

Operational Definitions of Extrinsic (Independent) Variables Examined: Credential Program Variables

Credit requirement: The total number of semester units (credits) required to complete a credential program. In the CSU this ranged from 32 to 56, as represented in university catalogs.

Program size: The number of candidates who completed a CSU credential program in one year. For the year of study, the CSU credentialed 6,667 elementary teachers, with a mean for campuses of 303 and a range from 66 to 570 teachers (CCTC, 2008).

K-12 School Variables

Socioeconomic status of students: The percentage of students in each graduate's class who qualify for free or reduced lunch.

Language status of students: The percentage of students in the graduate's class who are classified as English Language Learners.

Achievement level of the school: The decile ranking of the graduate's school of employment on the California Standardized Testing and Reporting (STAR) results.

Preparation of other teachers at the graduate's school: The percent of teachers in each school who were teaching on an "emergency permit" only.

Two Steps

This research was carried out in two steps. First, the effects of credit requirements and program size were examined by analyzing selected composite scores for respondent graduates and their employment supervisors on the *SEPTPP*. Second, the effect of the four student-related variables on selected composite *SEPTPP* scores for supervisors were examined.

Methods and Results

Study 1: Effect of Credential Program Variables on Composite Scores

Subjects: The subjects for the investigation of the effect of the two program variables on composite scores were graduates from a CSU elementary teacher credential program after one year of teaching (N=994) and their employment supervisors (N=242). Over 99% of the supervisors were school site principals (CTQ, 2009). According to survey results, 93% of employment supervisors reported visiting the first year teacher six or more times, and over 50% reported having greater than six conferences about teaching with the graduate. Graduates reported differently, with 78% reporting five or fewer observations by the principal, and 84% reporting five or fewer conferences about teaching with the principal. Either set of responses reveals multiple opportunities for supervisors to become familiar with the teacher's performance.

Dependent variables: The composites selected to examine the effect of the two credential program variables on ratings by principals represent the following important responsibilities of K-8 teachers: "How well prepared was the graduate to..."

- know and understand the subjects of the curriculum at the K-8 grade level?
- plan instruction and prepare classroom materials and activities for instruction?
- use an appropriate mix of effective teaching strategies in the classroom?
- meet the instructional needs of English Language Learners?
- understand child development, human learning, and the purposes of school?
- teach reading-language arts (K-8) according to state standards for the grade assigned?
- teach mathematics (K-8) according to state standards for the grade assigned?
- teach visual and performing arts according to state standards for the grade assigned?

The selected composites for the graduates represent their preparation to teach the following specific content areas: "How well prepared were you to..."

- teach reading-language arts (K-8) according to state standards for your grade?
- teach mathematics (K-8) according to state standards for the your grade?
- teach visual and performing arts according to state standards for your grade?

Correlation coefficients: Two-tailed Pearson correlations were used to investigate the relationship between the dependent variables and the independent credential program variables. Table 1 shows the correlation of principals' evaluation of the graduates' preparation with both the number of units required in each program and the number of credentials awarded for the year. Results showed no significant correlations between the independent and dependent variables. The correlations were small, ranging from -.066 to .062 for number of units and -.098 to .028 for the number candidates completing the program. As may be incidentally seen, the correlations among the dependent variables are all statistically and clinically significant. For example, the correlation between preparation to teach math and preparation to teach reading language arts was .942.

Toward a Culture of Evidence

Table 2 shows the correlation of the graduates' composite ratings with the two independent credential program variables. As with the principal evaluations, there were no significant correlations. The correlations ranged from .008 to .068 for program units and .019 to .046 for program size. Again, the relationship between the dependent variables were all statistically and clinically significant with the highest being .638, again between preparation to teach reading and preparation to teach mathematics.

Because the lack of a relationship between program length and preparation ratings seemed counter-intuitive, a follow-up examination of the data was conducted in which the CSU programs were segmented into two groups, those with 44 or fewer units required and those with 45 or more units. The F-ratios did not reach the statistically significant level for any of the dependent variables. These computations suggest that during the first year of teaching the reported levels of readiness by program completers to perform important responsibilities of K-8 teachers are not substantively related to the relative length of their credential preparation.

Table 1
Pearson Correlations of Principal Evaluation of Teacher's Preparation with the Number of Semester Hours Required in the Teacher's Credential Program, the Number of Credentials Issued, and the Inter-correlations Among the Evaluation Items (N=242)

Evaluation Item	1	2	3	4	5	6	7	8	# Sem. Hours	# of Cred.
Effectiveness of Preparation to:										
1. Understand Curriculum	__	.776*	.673*	.621*	.703*	.769*	.745*	.576*	-.013	-.091
2. Plan Instruct & Class Activities	.776*	__	.755*	.683*	.697*	.841*	.819*	.496*	-.066	-.039
3. Manage Class for Instruction	.673*	.755*	__	.622*	.639*	.716*	.688*	.453*	-.058	.024
4. Meet the Needs of ELL Students	.621*	.683*	.622*	__	.734*	.711*	.676*	.597*	-.045	-.044
5. Understand Growth & Develop.	.703*	.697*	.639*	.734*	__	.699*	.685*	.641*	.062	-.098
6. Teach Reading/ Language Arts (K-8)	.769*	.841*	.716*	.711*	.699*	__	.942*	.482*	-.012	-.042
7. Teach Mathematics (K-8)	.745*	.819*	.688*	.676*	.685*	.942*	__	.455*	-.06	-.031
8. Teach Visual-Perform Arts (K-8)	.576*	.496*	.453*	.597*	.641*	.482*	.455*	__	-.067	.028

* p<.01, two-tailed.

Study 2: Effect of K-12 School Characteristics on Composite Ratings

Subjects: The subjects for the investigation of the effect of the K-12 school variables on composite ratings were the employment supervisors of all elementary program completers since the initiation of *SEPTPP* who provided information for the four independent variables. A total of 19,050 supervisors responded to the survey over a 10-year period with responses for each independent variable ranging from 12,847 to 18,287.

Dependent variables: The composites selected to examine the effect on employment supervisors ratings of the characteristics of the schools in which the graduates were employed during their first year of teaching represent the following important responsibilities of K-8 teachers: “How well prepared was the graduate to...”

- meet the instructional needs of ELL learners?
- meet the instructional needs of learners from diverse backgrounds?
- meet the instructional needs of students with special learning needs?
- know about resources in school and community for students and families at-risk?
- communicate with parents or guardians of his/her students?
- teach reading language arts according to the *CA Content Standards in Reading*?
- teach math according to the *CA Content Standards in Mathematics*?
- use language so pupils at different levels of understand oral and written English?

Correlation coefficients: Table 3 shows the correlation between the four independent K-12 school variables and each of the eight dependent variables as well as the inter-correlations between the four independent factors. Each of the independent variables had a statistically significant correlation with at least one aspect of teacher preparation. The data, however, show that none of the correlations

Table 2
Pearson Correlations of Graduates’ Evaluation of Their Teacher Preparation with the Number of Units in Their Credential Program, the Number of Credentials Issued and the Intercorrelations Among the Evaluation Items (N = 994)

Evaluation Item	1	2	3	Semester Hours Required	Number of Credentials Issued
Effectiveness of Preparation to:					
1. Teach Reading Language Arts (K-8)	—	.638*	.520*	.047	.019
2. Teach Math (K-8)	.638*	—	.488*	.008	.020
3. Teach Visual Performing Arts (K-8)	.520*	.488*	—	.068	.046

* $p < .01$, two-tailed.

Toward a Culture of Evidence

Table 3
Correlation of CSU System-wide Principal Evaluation of Teachers' Preparation with Four Demographic Characteristics of the School

How well prepared was this teacher to begin each aspect of a teacher's job?	% of Students Eligible for Free or Reduced Lunch	% of Students Who Are English Language Learners	API Decile of School Last Year	% of Teachers in School with Emergency Permit
Number of Respondents	12,847	18,287	17,325	16,701
Meeting needs of ELL	.009	.004	.015	.024*
Meeting needs of diverse students	.009	-.001	.021	.017*
Meeting needs of students with special learning needs	.04*	.003	.06*	.02*
Knowing about resources for at-risk students	.02*	.004	.019*	.017*
Communicating with parents or guardians	.010	-.003	.04*	.007
Teaching standards-based reading/language arts	.045*	-.038*	.096*	.015
Teaching standards-based mathematics	.036	-.042*	.093*	.011
Use language so all pupils understand oral and written English?	.035	-.001	.062*	.018
Intercorrelations				
% Eligible Free/Reduced Lunch	—	.412*	.408*	.047*
% Students who are ELL	.412*	—	-.035*	-.028*
API Decile of School Last Year	.408*	-.035*	—	.003
% of Emergency Credentialed Teachers	.047*	-.028*	.003	—

Note. A total of 19,050 Supervisors responded to the survey. Different numbers of supervisors answered each of the questions thus the varying number of respondents.

* p < .05, two-tailed.

reached the .10 level. A correlation between two sets of scores reflects whether there is a consistent, predictable association between the scores. Researchers may interpret the magnitude and direction of the correlations as they stand, though most researchers prefer to square the correlation and use the resulting value to measure the strength of the relationship (Creswell, 2005). The coefficients of determination show that less than 1% of the variance of the supervisors' ratings is explained by any of the independent variables.

An examination of Table 3 shows that, as might be expected, a relationship between the SES at the schools, as measured by the percent eligible for free lunch, is strongly related to both the percentage of students who are ELL ($r=.412$) and the achievement level of the school ($r=.408$). The stereotype that low achieving schools are staffed with emergency permitted teachers was not demonstrated by this data. The correlations with the other independent variables were all less than .05, showing no clinical significance.

As was the case with the credential program variables, these computations taken together suggest that during the first year of teaching the reported levels of readiness by program completers to perform important responsibilities of teachers were not substantively related to conditions in the schools that are generally considered among educators, legislators, and the media and public to be an extreme challenge.

Discussion and Conclusions

Survey data is an important source of information for program assessment (Blanton et al., 2006; Cochran-Smith, 2009; Darling-Hammond, 2006). While student achievement data, process-product measures, and comparison to standards are also essential components of a comprehensive system to evaluate teacher preparation programs, the present research specifically examined the effect of certain extrinsic variables on principal and graduates' assessment of the graduates' university preparation program.

The impetus for such an examination lay with the lack of a culture of evidence at the campuses that have utilized the *SEPTPP* data. Cochran-Smith (2009) warned of a possible collision between the local culture of universities and the evidence used to examine for program quality. Gee (2007) and Phillips (2007) foreshadowed it, and Cochran-Smith and Zeichner (2005) overtly pointed out that there will be arguments about any program assessment data presented because of the complex web of variables involved in linking outcome measures to university program features.

The common experience across the CSU has been that, when faced with data reflecting less positively on the preparation program than they would like, faculty counter it with the rationale that the data reflect the external variables that were examined here. Study 1 addressed the most frequent argument, that more coursework would improve assessment data. California does not allow an undergraduate major in education so credentials are added on to degrees in other subjects and

Toward a Culture of Evidence

programs are also limited to preparation that can be completed in one year. Also stated frequently by faculty has been the belief that large preparation programs make students feel “distant” from instructors or professionally isolated. Neither the number of units required by the various programs nor the number of candidates completing a credential in a given year had any discernable effect on assessment data. The miniscule correlations between program length and the judged quality of preparation was surprising if not shocking. This data would seem to support the “point of diminishing returns” argument posited by Kaplan and Owings (2002) and should be seriously considered by all teacher education programs. The lack of effect from program size is less surprising, though it does counter those who advocate for small schools and small programs as a way to personalize and improve teacher preparation.

The relatively strong inter-correlations among the various aspects of teacher preparation that were rated indicates that teachers judged strongly prepared in one area are so judged in other areas. This may mean that the factors that make one a well-prepared teacher generalize across all areas of teaching or that strong preparation programs produce teachers that are strong across the board.

Study 2 addressed the assertion that teachers working in the most challenged schools, those with low SES, a high rate of ELLs, low achievement, and large numbers of emergency permitted teachers, will be judged less well prepared because they are teaching under more challenging conditions. It is acknowledged that new teachers are often placed in these schools because teachers with seniority flee these conditions when possible (Byrd-Blake et al., 2010). The results of Study 2, however, showed no clinically significant correlation between the principals’ evaluation of the CSU graduates’ preparation program and the characteristics of schools in which they taught during their first year. None of the variables reached even a minimal level of relationship. It is thus clear that principals’ judgment concerning the quality of a teacher’s preparation was not affected by the school characteristics that are typically thought of as indicating difficult teaching conditions.

The follow-up survey of university-based teacher preparation program graduates and employment supervisors conducted by the CSU is unprecedented. As predicted by the literature, however, some involved have been reluctant to accept this opportunity to utilize the culture of evidence so increasingly necessary in the field of teacher preparation. This research contributes knowledge to this critical area by addressing the extent to which the results of surveys assessing university-based teacher preparation are influenced by extrinsic variables over which a program has little or no control. The lack of relevant correlation found by these studies indicates that survey results can and should be used by programs to strengthen their preparation of future teachers without significant worry of contamination from the extrinsic variables examined.

Next Steps

Results of this study provide clear indications for future research to validate

the culture of evidence for program improvement in teacher education. Individual characteristics of the teachers being evaluated and effect of teacher ethnicity and background on supervisor evaluation should be studied. In addition, various pathways to becoming a teacher, including traditional campus-based preparation, professional development schools, residency programs, online programs, and programs where candidates are employed as teachers while completing their credential, should be compared both system-wide and within individual institutions.

A most important step in the extension of this study will be the triangulation of data including *SETPP* ratings, teacher performance assessments, and student achievement. While the statistical method known as “value added” is still being strongly questioned as a high stakes method (Baker et. al., 2010), student achievement is a factor that should be considered part of a rating of program efficacy. As stated, teacher work samples and performance assessments are useful for examining learning taking place in a classroom. Specific school achievement levels, by grade, subject, and subgroup is easily obtainable for all schools in California through the *Educational Results Partnership* website. While it may not be advised to use this data to evaluate individual teachers, it is facilitative to examine the effects a university-based program has on achievement at professional development schools or the locales where student teachers are placed. Finally, a comparison of traditional university-based teacher preparation with alternative programs such as *Teach for America*, paraprofessional teacher preparation programs, and for-profit institutions would provide valuable information for program improvement and would inform policymakers and future teachers about the validity of such alternative paths to teaching.

References

- Akmal, T., & Miller, D. (2003). Overcoming resistance to change: A case study of revision and renewal in a US secondary education teacher preparation program. *Teaching and Teacher Education, 19*, 409-421.
- Baker, E. L., Barton, P. E., Darling-Hammond, L., Haertel, E., Ladd, H. F., Linn, R. L., Ravitch, D., Rothstein, R., Shavelson, R. J., & Shepard, L. A. (2010, August 29). Problems with the use of student test scores to evaluate teachers. *EPI briefing paper #278*. Retrieved from the Economic Policy Institute website epi.3cdn.net/724cd9a1eb91c40ff0_hwm6ijj90.pdf
- Beare, P. (2009, February). The proof is in the pudding: Using data to document change. Paper presented at the American Association of Colleges for Teacher Education Annual Conference, Chicago, IL.
- Blair, J. (2000, October 25). ETS study links effective teaching methods to test-score gains. *Education Week, 20*(8), 24-25.
- Blanton, L. P., Sindelar, P. T., & Correa, V. I. (2006). Models and measures of beginning teacher quality. *The Journal of Special Education, 40*, 115-127.
- Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2008). Teacher preparation and student achievement. *Working Paper 14314*. Cambridge, MA: National Bureau of Economic Research. Retrieved from <http://www.nber.org/papers/w14314>

Toward a Culture of Evidence

- Byrd-Blake, M., Afolayan, M. O., Hunt, J. W., Fabunmi, M., Pryor, B. W., & Leander, R. (2010). Morale of teachers in high poverty schools: A post-NCLB mixed methods analysis. *Education and Urban Society, 42*, 450-472. doi: 10.1177/0013124510362340.
- Capraro, M. M., Capraro, R. M., & Helfeldt, J. (2010). Do different types of field experiences make a difference in teacher candidates' perceived level of competence? *Teacher Education Quarterly, 37*, 131-154.
- California Commission for Teacher Credentialing. (2008). *Teacher supply in California: A report to the legislature*. Sacramento, CA: Commisison on Teacher Credentialing. Retrieved from <http://www.ctc.ca.gov/reports/default.html>
- Center for Teacher Quality. (2006). *Preliminary work products: Systemwide evaluation of teacher preparation in the California State University*. Long Beach, CA: Office of the Chancellor, California State University.
- Center for Teacher Quality. (2009). *Systemwide evaluation of teacher preparation in the California State University*. Long Beach, CA: Office of the Chancellor, California State University.
- Cochran-Smith, M. (2003). Assessing assessment in teacher education. *Journal of Teacher Education, 54*, 187-191.
- Cochran-Smith, M. (2009). "Re-culturing" teacher education: Inquiry, evidence, and action. *Journal of Teacher Education, 60*, 458-468.
- Cochran-Smith, M., & Zeichner, K. (Eds.). (2005). *Studying teacher education: The report of the AERA Panel on Research and Teacher Education*. Washington, DC: American Educational Research Association.
- Cresswell, J. W. (2005). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (2nd ed.). Columbus, OH: Pearson.
- Darling-Hammond, L. (2002) Teacher quality and student achievement: A review of state policy analysis. *Education Policy Analysis Archives, 8*(1). Retrieved from <http://epaa.asu.edu/ojs/article/view/392>
- Darling-Hammond, L. (2006). Assessing teacher education. *Journal of Teacher Education, 57*, 120-138.
- Duncan, A. (2010). Teacher preparation: Reforming the uncertain profession. *Education Digest, 75*(5), 13-22.
- Gee, J. (2007). Reflections on assessment from a sociocultural-situated perspective. In P. Moss (Ed.), *Evidence and decision making, the 106th yearbook of the National Society for the Study of Education* (pp. 362-375). Malden, MA: Blackwell.
- Greenwald, R., & Hedges, L. V. (1996). The effect of school resources on student achievement. *Review of Educational Research, 66*, 361-397.
- Kaplan, L. S., & Owings, W. A. (2002). The politics of teacher quality: Implications and principals. *NASSP Bulletin, 86*(683), 22-42.
- Knapp, M., Copland, M., & Swinnerton, J. (2007). Understanding the promise and dynamics of data-informed leadership. In P. Moss (Ed.), *Evidence and decision making, the 106th yearbook of the National Society for the Study of Education* (pp. 74-104). Malden, MA: Blackwell.
- Levine, A. (2006). *Educating school teachers*. New York: The Education Schools Project.
- Louis, K. S. (2008). Creating and sustaining professional communities. In A. Blankstein, P. Houston, & R. Cole (Eds.), *Sustaining professional learning communities* (pp. 41-57). Thousand Oaks, CA: Corwin.
- Ludlow, L., Pedulla, J., Enterline, S., Cochran-Smith, M., Loftus, Salomon-Fernandez, &

- Mitescu, E. (2008). From students to teachers: Using surveys to build a culture of evidence and inquiry. *European Journal of Teacher Education, 31*, 319-337.
- Metzler, M. W., & Blankenship, B. T. (2008). Taking the next step: Connecting teacher education, research on teaching, and programme assessment. *Teaching and Teacher Education, 24*, 1098-1111.
- Moss, P. A. (2007). Reconstruction validity. *Educational Researcher, 36*, 470-476.
- Phillips, D. C. (2007). Adding complexity: Philosophical perspectives on the relationship between evidence and policy. In P. Moss (Ed.), *Evidence and decision making, the 106th yearbook of the National Society for the Study of Education* (pp. 376-402). Malden, MA: Blackwell.
- Torgerson, C., Macy, S., Beare, P., & Tanner, D. (2009). Fresno Assessment of Student Teachers: A teacher performance assessment that informs practice. *Issues in Teacher Education, 16*, 62-82.
- Wang, J., Spalding, E., Odell, S. J., Klecka, C. L., & Lin, E. (2010). Bold ideas for improving teacher education and teaching. *Journal of Teacher Education, 61*, 2-15.
- Whitehurst, G. J. (2002, March 5). *Scientifically based research on teacher quality: Research on teacher preparation and professional development*. Paper presented at the White House Conference on Preparing Tomorrow's Teacher, Washington, DC. Retrieved from http://scholar.google.com/scholar?q=whitehurst+2002&hl=en&btnG=Search&as_sdt=2001
- Wilson, S., Floden, R., & Ferrini-Mundy, J. (2001). *Teacher preparation research: Current knowledge gaps, and recommendations*. Seattle, WA: Center for the Study of Teaching and Policy, University of Washington. Retrieved from <http://depts.washington.edu/ctpmail/PDFs/TeacherPrep-WFFM-02-2001.pdf>
- Zeichner, K. M. (1999). The new scholarship in teacher education. *Educational Researcher, 28*(9), 4-15.