

**Assessing CSU Teacher Preparation in Relation to K-12 Student Learning:
Preliminary Findings Based on Incomplete Evidence**



William Wilson
Assistant Vice-Chancellor

David Wright
Principal Investigator

Nohoon Kwak
Evaluation Statistician

**CSU Center for Teacher Quality
Office of the Chancellor
May 1, 2006**

First Evaluation Question
With Preliminary Findings Based on Incomplete Evidence

Question: To What Extent Do Variations in Teachers and Their Preparation Account for Variability in Student Learning as Assessed by California Standardized Achievement Tests?

CSU Partnerships with California School Districts. During 2004-05 and 2005-06, the *Systemwide Evaluation of Teacher Preparation* examined teacher preparation in relation to *learning gains by the K-12 students* of CSU teachers. To address the *First Evaluation Question*, CSU campuses and the Chancellor's Office (CO) are entering into partnerships with large urban school districts in Los Angeles, Long Beach, San Diego, San Jose, Fresno and Sacramento. In each partnership, the CO and CSU campuses work with district data officers to exchange available data regarding (a) teachers and their preparation, and (b) the demographic, socio-economic and academic records of the teachers' students. Before the partners exchange data in these partnerships, agreements are reached regarding data security, protection of teacher and student privacy, evaluation purposes and evaluation methodology.

Preferred Statistical Test. To address the *First Evaluation Question* statistically, the CSU utilizes *hierarchical linear modeling*, a special type of *multiple regression* that accounts for the fact that students in a classroom influence each other. Unlike traditional *regression analysis*, *HLM* does not assume that learners in a class are independent of each other. Instead, evaluators who use *HLM* build a multi-level conceptual model in which different independent variables are at *different levels of a "nested design."* For this evaluation, the Chancellor's staff designed a hierarchical model in which individual student variables enter the equation at Level One and variables that describe teachers, classrooms or schools enter at Level Two. *HLM* indicates the extent to which selected *independent variables*, including teacher preparation variables, account for variability in *dependent variables* such as student test scores, while controlling for the model's other independent variables. *HLM* models can include as many independent variables -- including student demographics -- as are available in coded data. Education researchers generally consider *HLM* to be the most suitable statistical test for evaluations of K-12 student learning and the sources of that learning.

Methodology of the Preliminary Analysis. The CO examined scores earned on SAT-9 and CAT-6 examinations in 2002 and 2003 by 21,000 fourth-grade students and 16,000 fifth-grade students in one large district. The analysis was an example of *value-added assessment* because each student's record included a pretest score (earned at the conclusion of the prior school year) and a posttest score (earned at the conclusion of the year of assessed instruction). Pretest scores were treated as *covariates* in the analysis of posttest scores. Students with incomplete pre-post records were not included. *This work is "preliminary" because (1) the CO is still analyzing scores on other tests and in other grades and subjects, (2) the CSU is requesting additional data from this district and from other large districts in California, and (3) the CO would like to work with CSU campuses to enrich the teacher preparation data for subsequent analyses.* This paper provides an early glimpse at a multi-year effort. Comments by CSU campuses are invited. Errors of fact, method or interpretation are the CO's responsibility.

Variables in the CSU's Nested Design. The preliminary analysis focused on *pupils' posttest scores in reading, language and mathematics*. In the CO's multi-level model, *each pupil's record* at Level One included the student's subject-specific pretest score, English language proficiency level, eligibility for free/reduced school meals, disability status, gender and ethnicity. *Each teacher's record* at Level Two included the source of preparation, the length of prior teaching experience in the school district, and the teacher's reported ethnicity and gender. Prior to completing further analyses, *it would be advantageous for CSU to enrich the teacher preparation data in ways that are suggested below*.

Two Distinct Statistical Models. The CO used two distinct statistical models to examine the extent to which teacher variables accounted for variability in students' posttest scores. First, the CO examined posttest scores by themselves, *without taking account of student variables* in the data file (called the Unconditional Model). Then the student variables were entered into a second equation (with listwise deletion) in which the impact of teacher variables was assessed *while student factors were taken into account* (called the Model with Level-I Covariates).

Teacher Effects in the Two Models. Initial analysis of the first model (Unconditional) indicated that *21 percent to 25 percent of variance in student learning was attributable to variations in teacher preparation*. Variability in teachers and their preparation had the *greatest impact* on mathematics learning scores in the fourth (23%) and fifth (25%) grades, and the *least impact* on language learning scores in the fourth (22%) and fifth (21%) grades. Initial analysis of the second model (Level-I Covariates) indicated that six percent to 13 percent of learning variance was attributable to teacher-preparation variables, with the *greatest impact* in mathematics (13% and 11% at the two levels) and the *least effect* on language learning (7% and 6% respectively). *These learning effects were primarily attributable to varied teacher preparation*. Variability in teachers' classroom experience (from one to four years) was *not* strongly related to student outcomes, especially in math.

Effects of Student Independent Variables. In the Model with Level One Covariates, initial analysis suggested that students' pretest scores were *the most influential factors* in accounting for variation in posttest scores by fourth- and fifth-grade students. Other significant factors were students' English language proficiency levels, eligibility for free/reduced school meals, disability status, gender and ethnicity.

Preliminary Interpretation by the Chancellor's Office. In a preliminary interpretation, the CO considers the findings to underscore the *critical importance of teacher preparation*. In this analysis, the CO intentionally focused on variations in teacher preparation while holding all other learning factors constant in its statistical methodology. By doing so, the CO showed that *substantial amounts of student learning in elementary mathematics, reading and language are linked to variations in the preparation of teachers*. This finding is not inconsistent with prior research showing the importance of student demographic factors such as family poverty and students' levels of English proficiency. These factors *are* significant. The CO's design, however, indicates that *teacher preparation is also significant*, even when student factors are taken into account. Indeed, *teacher preparation is the only variable in the analysis that can become more effective* through educational research, innovation and evaluation. California schools cannot change the extensive variability that characterizes the personal backgrounds of California students. The California State University and its counterparts in the UC and independent segments can, however, continue to modify teacher preparation to maximize student learning in K-12 schools. The analysis confirms that improvements in teacher preparation, such as the CSU Deans' *Excellence in Teacher Preparation Initiative*, are potentially effective strategies in strengthening the education of California's sons and daughters. This finding is likely to be intensified if CSU campuses were to enrich the teacher preparation data they provide to the Chancellor's Office. The CO looks forward to further collaboration with CSU campuses and several California districts in an ongoing effort to examine the effects of teacher preparation on student learning in recent years, and to strengthen those effects in the future.

Second Evaluation Question

With a Preliminary Analysis for Review and Discussion by CSU Deans of Education

Question: Compared with Universities Outside the CSU System, What is the Relative Level of CSU Effectiveness in Fostering K-12 Learning through the Preparation of New Teachers?

In addressing this question, the Chancellor's Office (CO) examined *the same teacher-student data* that were analyzed on pages 1-2. These data, provided by a large district in California, included the teachers of 21,000 grade-four students and 16,000 grade-five students. Because of differences in the two evaluation questions, the CO used *distinct methodologies* in order to examine their results in exploratory ways.

Groups of Students and Teachers. With the school district's assistance, the CO split the fourth- and fifth-grade students into two groups, respectively, based on the *sources of their teachers' preparation*. In both grades, the first group consisted of students taught by 196 teachers who *had recently completed CSU programs of professional preparation*. The second group consisted of students of 1,428 other teachers who had *not recently completed CSU programs*. The 196 recent CSU teachers in the first group had completed all credential requirements at CSU campuses and had earned full, unrestricted California credentials in 2001 or 2002, either one or two years prior to dates when their students earned the test scores being analyzed here. Some of the teachers in the second group *may* have earned credentials at CSU campuses, but *they did so before 1999*. Having completed their *professional preparation at distinct universities or during distinct periods of time*, the two groups of teachers were *matched with each other in teaching experience*: Members of both groups had from one to four years of experience as teachers in the school district when their students earned the reported test scores in 2002 and 2003.

Preferred Statistical Test. To address the *Second Evaluation Question* statistically, the CO *contrasted mean pretest scores earned by the two student groups with their mean posttest scores*. This methodology relies on tests of statistical significance and mean-score effect sizes. The design provides for a *head-to-head comparison* between the effectiveness of the two groups of teachers. It also highlighted differences that may exist between the two groups of students in their levels of prior mathematics, reading and language learning. In this second analysis, the CO examined the extent to which differences in the two student groups' scores could be attributed to the preparation of their teachers (which was the principal difference between the teacher groups) OR to random variability within the two groups of students (e.g. factors not being contrasted by separating teachers recently prepared by the CSU from teachers not recently prepared by CSU).

Comparing the Aggregated Effectiveness of the Two Groups of Teachers. Graphs on pages six through eight illustrate a prototype methodology for investigating the *Second Evaluation Question*. *Figure One* portrays the performances of students and teachers in terms of **reading** scale scores on pretest (SAT-9) and posttest (CAT-6) examinations in 2002 and 2003. *Figures Two* and *Three* provide similar portrayals of **language** and **mathematics** scale scores by the same students taught by the same teacher groups. The numbers of participants included in this analysis are slightly greater than the *Ns* cited earlier because the CO did not remove students from this analysis if their records were incomplete. Mean values are based on scores earned by students who actually took each test. The analysis is *not* an example of *value-added assessment* because it does *not* depend on longitudinal tracking of each student's pre-post sequence of scores.

Preliminary Interpretation of Aggregated Effectiveness of Two Teacher Groups. Figures 1-3 show *strong relationships between pretest and posttest scores*, confirming a major finding in the *HLM* results. The graphs also compare the two groups of teachers and pupils in terms of (1) the students' learning status prior to instruction by the teachers and (2) their relative positions following a year of assessed instruction. *Figure One* suggests that CSU teachers produced instructional results in grade five reading that were *higher than those of the comparison teachers and students* ($p < .05$), when the two groups of students were at equivalent levels of *prior learning*. *Figure Two* suggests that CSU teachers "leveled the field" for their students in grade four language instruction although their students were "behind" their counterparts prior to instruction ($p < .001$). *Figure Three* indicates that mathematics instruction by the two teacher groups in grades four and five *did not measurably alter the relative positions of the two groups of students* in their pre-instruction and post-instruction levels of proficiency.

Relative Importance of the Aggregated Group Findings. The three findings outlined above are based on the fact that mean variations between the two groups of students were relatively large compared with similar variations among individual students within the groups. These statistical differences should not be assigned too much importance, however, because the two groups of students were relatively large. Effect sizes, which are measured in terms of standard deviations from the means, were clearly modest. *How much importance should the CSU assign to the findings? Given the tendency of state and federal policymakers to value test-score gains, CSU should not turn away from learning differences that are statistically attributable to teacher preparation.* Still, it would be a mistake to assign too much importance to an analysis that was undertaken primarily to explore the viability of an evaluation design, not to confirm or disconfirm a prior hypothesis.

Examining the Effects of Teacher Preparation on Important Groups of Underserved Students in California. The *HLM* analysis on pages 1-2 examined demographic factors that cause some groups of students to be served ineffectively at all levels of education. The *Second Evaluation Question* provides an opportunity to examine the academic status of under-served student groups before and after they are taught by the two groups of teachers. To pursue this opportunity, the CO examined pretest and posttest results for the two groups of teachers among grade-four and grade-five students who were in distinct groups on the basis of family income, prior English proficiency, and disability status. The CO also examined similar results for boys and girls, and for members of five ethnic groups. Separate analyses of reading, language and math scores on pretests and posttests in grades four and five yielded a total of 60 statistical analyses for *the purpose of discerning how well CSU teachers and other teachers serve groups of students who have historically been under-served in education.*

Analysis and Interpretation of Results Pertaining to Underserved Students in California. Space limitations preclude a thorough explication of all 60 analyses that the CO completed. Overall, none of the 60 analyses indicated that student groups were disadvantaged by recently-prepared CSU instruction. Some of the 60 analyses indicated that recently-prepared CSU instruction was more effective than instruction by other teachers. In grade five, for example, low-income students made greater gains in reading, language and math while being taught by recent CSU graduates (pupil N = 2,100) compared with their low-income classmates taught by other teachers (pupil N = 16,000). The same patterns were found when the analysis focused on English learners. Hispanic and African-American students were advantaged by recently-prepared CSU teachers at both grade levels. The two groups of teachers were equally effective in teaching students with disabilities. The CO plans to report these findings in greater detail. *Concurrently, it would be valuable and important for CSU campuses to identify individual graduates from 2002-03 and 2003-04 who completed specific pathways or cohorted programs of teacher preparation. With enriched campus data, a Dean will receive campus-specific findings, program-specific findings, and pathway-specific in which the K-12 student learning gains are reported alongside the evaluation judgments of recent CSU graduates and their employment supervisors.*

New Evaluation Questions Being Discussed by CSU Campuses and the Chancellor's Office

Based on the preliminary analysis of K-12 pupil data, the CO and CSU campuses drafted the following *new Evaluation Questions* to encourage critical review and extensive CSU discussion about the prospect of linking K-12 student learning to the preparation of CSU and other teachers.

| | | |
|---|--|---|
| <p style="text-align: center;">First New CSU Evaluation Question:</p> <p style="text-align: center;">On the Importance of Teacher Preparation</p> <p><i><u>How much of student learning is associated with the following teacher preparation and certification variables when we take student and teacher characteristics into account?</u></i></p> <ul style="list-style-type: none"> ● <i>The institutions/organizations where teachers complete their professional preparation (universities/districts).</i> ● <i>Significant variations in fieldwork by prospective teachers during preparation.</i> ● <i>Variations in professional/state/national standards that govern teacher preparation.</i> ● <i>Levels of California state certification that reflect different levels of preparation.</i> ● <i>State policy variations that governed the preparation of teachers during different periods of time.</i> ● <i>Intervals of time that elapse between the assumption of K-12 teaching responsibilities and the completion of accredited professional preparation by individual teachers.</i> | <p style="text-align: center;">Second CSU Evaluation Question:</p> <p style="text-align: center;">Overall CSU Effectiveness</p> <p><i><u>In preparing new teachers, what is the CSU system's overall level of effectiveness relative to other institutions in the following series of analyses that control for teacher and student variables?</u></i></p> <ul style="list-style-type: none"> ● <i>Analysis of the shape of learning distributions among K-12 students whose teachers were professionally and recently prepared by CSU, compared with other K-12 students.</i> ● <i>Analysis of statistical significance and effect sizes in mean levels of learning among the students of CSU teachers compared with mean levels among other students.</i> ● <i>Analysis of mean levels of student achievement for CSU-prepared teachers, teachers from California's other institutions, non-California institutions, and California school districts.</i> | <p style="text-align: center;">Third New CSU Evaluation Question:</p> <p style="text-align: center;">Features of Most Effective CSU Programs</p> <p><i><u>Which CSU campuses and programs are most effective in promoting the academic achievement of K-12 students when we also take important student and teacher variables into account?</u></i></p> <ul style="list-style-type: none"> ● <i>Analysis of <u>effective</u> CSU campuses and programs (ones that are significantly associated with K-12 pupil learning) to determine their distinctive programmatic features or attributes.</i> ● <i>Analysis of the relative strength of statistical links between K-12 student learning and the identifiable features and attributes of the most <u>effective</u> CSU programs for new teachers.</i> ● <i>Analysis of the potential feasibility of describing and disseminating the features and attributes of most effective programs in order to encourage others to replicate them.</i> ● <i>Analysis of the potential feasibility of designing and conducting quasi-experiments to explore the effects of preparation that is designed primarily to improve K-12 learning.</i> |
| <p>The CSU Chancellor's Office would welcome campus comments and recommendations about these new draft evaluation questions.</p> | | |