Does Teacher Professional Development Affect Content and Pedagogical Knowledge: How Much and for How Long?

CSE Technical Report 711

Pete Goldschmidt
CRESST/University of California, Los Angeles

Geoffrey Phelps
University of Michigan

April 2007

National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)
Center for the Study of Evaluation (CSE)
Graduate School of Education and Information Studies
University of California, Los Angeles
GSEIS Bldg., Box 951522
Los Angeles, CA 90095-1522
(310) 206-1532

The work reported herein was supported under the Educational Research and Development Centers Program, PR/Award Number R305B60002, as administered by the Office of Educational Research and Improvement, U.S. Department of Education.

DOES TEACHER PROFESSIONAL DEVELOPMENT AFFECT CONTENT AND PEDAGOGICAL KNOWLEDGE: HOW MUCH AND FOR HOW LONG?

Pete Goldschmidt
University of California Los Angeles

Geoffrey Phelps
University of Michigan

Abstract

We examine the impact of teacher professional development on knowledge growth and subsequent knowledge retention. Specifically we use English Language Arts teacher content and pedagogy assessments to determine whether the California Professional Development Institutes significantly improve teacher content knowledge and whether teachers retain that knowledge six months after the institutes are completed. The results indicate that teachers vary significantly in pre-institute knowledge on the four assessed domains, demonstrate significant knowledge growth, but only retain about one half of what was gained during the institute. Further, pre-existing knowledge gaps are not systematically reduced and teacher perceptions of institute quality are not related to knowledge growth and knowledge retention.

Current No Child Left Behind (NCLB) legislation requires that all students are taught by highly-qualified teachers (NCLB, 2001). The logic of NCLB and related state and district policies is straightforward; to make adequate progress towards proficiency in the core subjects of English language arts and mathematics, students need to receive instruction from teachers who are well-prepared to teach these subjects. Teacher professional development is widely viewed as the most promising intervention for improving teacher quality in U.S. public schools. Nearly every state and district provides inducements for teachers to participate in professional development and these are increasingly supplemented by federal programs such as Reading First. Districts often add requirements on top of these, either in the form of mandatory programs for all instructional staff, or by requiring time investment beyond state requirements. The

ubiquity of professional development is apparent in teachers' own reports of their professional learning activities. Data compiled by the National Center for Educational Statistics (NCES) shows that in 1999-2000, 99% of teachers surveyed reported participating in professional development activities within the past year.

While resources are being poured into professional development, evidence for the effectiveness of these programs is uneven. Practical and methodological complications often frustrate efforts to establish clear relations among professional development, teacher learning, instructional improvement, and student outcomes. Studies that seek to examine these relations are typically small-scale involving just a few teachers or at most a few schools. When professional development is examined in large-scale studies typically only the relation of program and student achievement are examined (Garet, Porter, Desimone, Birman, & Yoon, 2001). When teacher outcomes are included in large-scale studies, the measures are often based on teacher opinions and self-evaluations. Teachers are asked, for example, if they liked the professional development or if they learned something new. Few large-scale studies directly measure teacher knowledge and its growth or seek to examine the effects of professional development on teacher learning (Carlisle, Correnti, Phelps, & Zeng, 2006).

We report results from a study of the California Professional Development Institutes (CPDI)—a large-scale professional development initiative designed to improve the content knowledge of teachers in the elementary subjects of reading and mathematics. The analyses presented here make use of an innovative measure of teachers' knowledge about reading. Teacher knowledge is measured in the summer pre- and post-CPDI, with a final follow-up measure the following spring. The longitudinal design and use of equated measures provides a unique opportunity to investigate whether changes in teacher knowledge can be attributed to in-service professional development and if these changes are sustainable over time. The results indicate that in-service professional development can affect changes in the knowledge teachers use in elementary reading instruction, but that knowledge gains erode after teachers return to the classroom.

Background on the California Professional Development Institutes

In 2000, the state of California launched the California Professional Development Institutes (CPDI), a statewide initiative designed to improve teachers' content-based knowledge and skills with the goal of ensuring that all students learn to read by the 3rd grade. The CPDI grew out of what was previously the Governor's Professional

Development Institute. However, by design the CPDI differed from previous teacher development initiatives in a number of important ways. Perhaps most obvious was the unprecedented scope of the CPDI initiative. While the focus of the institutes was initially on early career teachers, especially those working in traditionally disadvantaged schools, the eventual goal was to involve *every* elementary English Language Arts teacher in California.[1] The CPDI also stood out from prior professional development initiatives in its explicit goal of improving the subject matter knowledge of teachers in areas defined by the state standards (Griffin, Aguirre-Muñoz, Miyoshi, Roberson, & Amabisca, 2001).

The CPID was comprised of four structural models. These were the California Reading and Literature Project (CRLP), the California Writing Project (CWP), materials-based institutes,[2] and faculty-designed institutes (Griffin et al., 2001). A majority of the institutes were either CRLP or CWP. One major difference between CRLP and CWP was that while both intended to cover reading and writing as part of the CPDI, CRLP initially focused on reading and literature while CWP focused on writing (Griffin et al., 2001).

Although specific sites used different instructional models, there were a number of features common among all CPDIs.[3] All sites included an initial 40-hour summer institute, 40 hours of follow-up professional development during the school year, and 40 hours of team meetings. All CPDIs encouraged teacher teams and administrators from participating schools to engage in collaborate problem-solving strategies to enhance school-wide improvement efforts (Griffin et al., 2001).

The over-arching premise of the CPDI was that research-based content and pedagogic strategies must be utilized to provide teachers with opportunities to develop knowledge, skills, and motivation to change their practices (Griffin et al., 2001). The CPDIs reflected current research on effective professional development (Darling-Hammond, 1999) utilizing participatory formats that relied on teacher expertise to lead a majority of activities. Overall, about 60% of CPDI activities were participatory (Griffin et al., 2001). Approximately 15% of CPDI time was allocated to teachers practicing specific instructional activities (Griffin et al., 2001).

---

[1] Similar goals were cited for creating Mathematics CPDIs, but we examine only ELA CPDIs here.

[2] These were district-specific institutes providing training for specific district curricular materials. The majority of these consisted of the Open Court reading program.

[3] Some of these features were required in the request for proposals (RFP).

The content coverage of the CPDIs was divided among word analysis, comprehension, and writing. About one quarter of the time[4] was spent on deepening teachers' understanding of reading research as well as specific aspects of teaching reading, such as the six stages of learning to read (Griffin et al., 2001). A third of the activities included phonemic awareness, phonics, decoding/word recognition, and vocabulary building. Coverage of writing tended to be less comprehensive as between about 10% and 25% of activities addressed comprehension, communication, and other aspects of writing (Griffin et al., 2001). Given the CPDI emphasis on instructional practices, less time was devoted to assessment and diagnostic strategies (Griffin et al., 2001). Spanish language issues were only given cursory coverage in the CPDIs (Griffin, et al., 2003b).

Previous analyses of the CPDIs examined the impact on teacher attitudes and practice (Griffin et al., 2001, 2003a, 2003b). There were some positive changes in self-reported attitudes and practices. Overall, teachers were satisfied with their experiences, and the content and organization of the CPDIs; although, less so for the 40 hour follow-up (Griffin et al., 2003b). Teachers reported that the CPDIs were helpful in guiding and improving both classroom instructional and assessment practices, but more so in reading than writing (Griffin et al., 2003b).

Background on Teacher Knowledge

By design the California Professional Development Institutes set out to improve teacher knowledge, in particular teacher knowledge as it relates to teaching word reading and reading comprehension. On the surface this is a straightforward goal. It is a matter of common sense that teachers need to know the subjects that they teach. To be effective in teaching science or mathematics, teachers themselves must understand the content that their students are learning. Less clear, however, is what exactly counts as content knowledge in the area of elementary reading. After all, reading is typically thought of as something that individuals *do*, not something that individuals *know*. In order to better understand what there is for teachers to learn about reading itself in professional development programs such as the CPDI, it is useful to briefly review current thinking about teacher subject matter knowledge in general and, in particular, in the area of reading.

[4] Content coverage is not mutually exclusive.

Teacher educators, policy makers, and others concerned with teacher quality and preparation have long recognized the importance of teachers having a basic mastery of content. More recently, concern over subject matter preparation is incorporated in state and federal policies. The NCLB Act of 2001 requires that teachers pass a state test demonstrating subject matter knowledge in reading, writing, mathematics, and other core subject areas (NCLB, 2001). The assumption is that teachers need to know content as it is taught and learned in college and university classes. Better prepared teachers have taken *more* classes in mathematics or science. In the area of reading it would follow that better prepared teachers are themselves better readers. These are after, all sensible ideas. They are, however, just a start on the subject matter knowledge needed to teach effectively.

Evidence is mounting that teaching a subject requires content knowledge that goes substantially beyond what is typically taught and learned in college and university classes. This special form of content knowledge is most commonly referred to as pedagogical content knowledge or simply PCK (Ball, 1988, 1991; Ball, Lubienski, & Mewborn, 2001; Gess-Newsome & Lederman, 1995; Grossman, 1990, 1991; Leinhardt & Smith, 1985; Magnusson, Krajcik, & Borko, 1999; Shulman, 1986, 1987; Wilson, Shulman, & Richert, 1987; Wilson & Wineburg, 1988; Wineburg & Wilson, 1991). The term PCK was original coined by Lee Shulman and was defined as "the most useful forms of representation … the most powerful analogies, illustrations, examples, explanations, and demonstrations—in a word, the most useful ways of representing and formulating the subject that make it comprehensible to others…. Pedagogical content knowledge also includes an understanding of what makes the learning of specific topics easy or difficult" (Shulman, 1986, p. 7).

Teaching reading, like other subjects, requires knowledge that goes substantially beyond just being a good reader. In addition to having strong reading skills, well-prepared reading teachers also need to develop deep knowledge of language and text; for example, reading teachers need to understand reading in ways that help them decipher the stumbling attempts of a beginning student or to select appropriate words or text for students of different ability levels. Emerging arguments and evidence suggest that it is this knowledge about reading itself is that poorly understood by literate adults who do not teach children to read (Brady & Moats, 1997; McCutchen, Abbott et al., 2002; McCutchen & Berninger, 1999; McCutchen, Harry, & Cox, 2002; Moats, 1994, 2000; Moats & Lyon, 1996; National Board for Professional Teaching Standards [NBPTS], 2001; Phelps & Schilling, 2004; Wong-Fillmore & Snow, 2002).

Teacher Knowledge Measures and Descriptive Results

At the time of the CPDI, reading researchers had only just begun to turn their attention to studying teacher content knowledge. The few survey instruments that were available assessed teacher knowledge of English language word structure with a focus on early primary grade topics (e.g., phonemic, letter-sound relationships, spelling patterns; Bos, Mather, & Dickson, 2001; Cunningham, Perry, Stanovich, Stanovich, & Chappell, 2001; Moats, 1994). None of these instruments were suitable for measuring the knowledge addressed in the CPDI institutes. At the same time that the CPDI research team was searching for a suitable measure, the Study of Instructional Improvement, a large-scale mutli-method study of whole school improvement programs, was in the final stages of developing a large pool of multi-choice survey items designed to study elementary teacher knowledge. The CPDI research team and the Study of Instructional Improvement agreed to collaborate. The CPDI institutes served as a pilot site for field testing the newly written items and the pilot assessments were designed to yield multiple equated measures suitable for studying teacher growth.

The Content Knowledge for Teaching Reading (CKTR) items developed by the Study of Instructional Improvement were written to assess teacher knowledge in the two broad topic areas of comprehension (e.g., morphology, vocabulary, comprehension strategies and questions, genre, fluency, and other topics related to comprehending the meaning of words and text of word reading) and word analysis (e.g., phonemic awareness, letter sound relationships, word frequency, and other topics related to the reading and decoding of words and their print and sound elements).

In addition to these broad topic domains, items were also written to represent three cross-cutting distinctions in how teachers use knowledge in instruction. Knowledge of content and teaching (KCT) items require respondents to use knowledge of reading to develop or choose teaching actions or moves. This may require using knowledge of content to determine what to say to a student struggling over a difficult word or what type of task to give a student to help her comprehend a challenging passage. Items that tap knowledge of content and students (KCS) require respondents to use knowledge of reading to decipher and interpret students' products or work. This may entail recognizing typical student errors or approaches to reading. Items in this category do not require making teaching decisions. Content Knowledge (CK) items require respondents to use knowledge of reading in the context of teaching situations. While these items may require a specialized knowledge of content beyond what most

adult readers typically understand, they do not require knowledge of students or teaching. Finally, a small set of items was written that focused the special knowledge of English and Spanish used when teaching English language learners.

A total of 257[5] unique items were administered across three forms to teachers participating in the summer CPDI institutes. Each of the three forms included roughly the same number of comprehension and word reading items. Within the larger domain of comprehension, items were included on each form that focused on CK, KCS, and KCT. In the domain of word analysis, items were only available for the domains of CK and KCS. Forms were randomly assigned by institute to pre- and post-administration. Participants in any given institute, therefore, received either form A, B, or C as a pre-institute assessment with one of the two remaining forms randomly assigned as a post-institute assessment. Preliminary results from the pre- and post-institute assessments guided item selection for a single final form to be administered to all participants who had previously completed the pre- and post-assessment forms. Selections for the final form were made from within each domain based on item discrimination, item difficulty, and other considerations such as the quality of item design and the relevance of the item to elementary instruction.

The final assessment included 78 questions representing each of the five major domains and was completed by 599 participants. Full information exploratory and confirmatory factor analyses were conducted to determine the structure of the CKTR items (Thissen & Wainer, 2001). Three unique factors were identified. These factors closely mapped onto the five theoretical domains discussed above. In both comprehension and word analysis no distinction was found between the CK and KCS items. Since the KCS items did not form a separate factor, they are included with the CK items and are referred to henceforth as content knowledge items. The KCT items in comprehension, however, were distinct from both the other comprehension items and from the word analysis items. The results from the factor analysis are reported in detail in Phelps and Schilling, 2004. The Spanish language items were not included in the factor analysis. However, since these items require knowledge of Spanish, they most likely form their own unique knowledge domain.

The CKTR items on each of the pre- and post- institute assessments and the final assessment were grouped according to the factor analysis results into four scales:

[5] The three forms shared a large number of common items. This provided a basis for assessing the relative difficulty of the three forms and assisted in item selection for the final assessment. When common items are included, a total of 429 items were administered across the three forms.

Comprehension /Content Knowledge (CMP/CK); Comprehension /Knowledge of Content and Teaching (CMP/KCT); Word Analysis Content Knowledge (WA/CK); Spanish language (SPAN).

Two parameter Item Response Theory (IRT) models were used to examine item properties, develop and equate scales, and score participants (Hambleton, Swaminathan, & Rogers, 1991). The software program BILOG was used for all IRT scale estimations (Mislevy & Bock, 1997). In order to achieve a larger sample size and to improve item parameter estimates, data from pre- and post-institute administrations for each form were combined (Form A=798 , Form B=1000, Form C=1170).[6] Variation in the total number of participants completing each of the three forms is due to differences in the size of institutes to which each form was assigned. To study teacher growth from pre-test, to post-test, to the final assessment, scale scores for each domain must be calibrated or put on the same ability metric. The three pre- and post-test forms and the final assessment were calibrated using an anchor test design. Specifically, each of the three pre- and post-assessment forms shared a set of common or linking items with the final assessment (Table 1).

The difficulty parameters for common items on different test forms are used to estimate scale (α) and location (β) constants that can then be used to place each of the different assessments on the same ability metric (Hambleton et al., 1991). Scale and location constants are estimated such that

$$\alpha = \frac{s_{Yc}}{s_{Xc}}$$

and

$$\beta = \bar{b}_{Yc} - \alpha\ \bar{b}_{Xc} + \beta$$

[6]A potential concern with combining pre- and post- data is that participants will learn particular content from the institutes and this will have differential impact on particular items. A parallel analysis was conducted using just the pre-institute data for forms A, B, and C. Final scores for estimations using pre and post-institute data and just pre-institute data had extremely high correlations. All the comprehension and word analysis scales had correlations above .97 and the Spanish language scale had correlations above .94.

where $\bar{b}_{Yc}$ and s_{Yc} are the mean and standard deviation of estimates for the difficulties of the common items in test Y and $\bar{b}_{Xc}$ and s_{Xc} are the mean and standard deviation of the estimates of the difficulties of the common items in test X. The parameter estimates for test X are placed on the same scale metric as test Y using

$$b_Y^* = \alpha \; b_X + \beta$$

$$a_Y^* = \frac{a_X}{\alpha}$$

where a_Y^* and b_X^* are the difficulty and discrimination values respectively of items on test Y and X.

Since this equating method assumes a linear relationship, it is important to first establish that all common or linking items fit a linear equation and include items that represent a range of difficulties. To identify problematic items, all difficulties for linking items were plotted and misfits deleted (for details see Phelps, 2003). The equating constants appear to be well estimated for all scales with the exception of Spanish language. These scales include a limited number of common items and these items are all easy with a very limited difficulty range. Therefore, the equated scores of the pre- and post-institute assessments for Spanish are likely to be somewhat unreliable. Bilog was used to re-estimate item parameters for each scale on forms A, B, and C, placing each on same scale metric as the scales on the final assessment.

Table 1

Content Knowledge for Teaching Scale Measures

Scale	Number of items	Number of linking items	IRT reliability	Test information curve maximum
Form A				
CMP/CK	40	13	.82	-1.1
CMP/KCT	25	9	.73	-1.9
WA/CK	54	26	.90	-2.1
SPAN	9	3	.68	-0.9
Form B				
CMP/CK	59	19	.88	-1.4
CMP/KCT	14	9	.74	0.4
WA/CK	64	35	.91	-1.9
SPAN	11	4	.66	-0.8
Form C				
CMP/CK	65	18	.89	-1.8
CMP/KCT	21	12	.78	1.2
WA/CK	56	26	.90	-1.9
SPAN	11	3	.71	-0.9
Final Post Test				
CMP/CK	24		.79	-2.1
CMP/KCT	18		.76	0.6
WA/CK	35		.86	-1.9
SPAN	5		.75	-1.0

The reliabilities of the scores range from a low of 0.66 for the 11 item Spanish scale to a high of 0.91 for the 64 item WA/CK scale. With the exception of Spanish language all scales have moderate to high reliabilities. Nearly all of the scales are providing maximum information for participants between 1 and 2 standard deviations below the mean. Only the CMP/KCT scale is providing maximum information for participants with abilities above the mean. This indicates that overall the teacher knowledge measures will give very reliable estimates for participants with less knowledge. Participants with high levels of knowledge will be measured less reliably. Finally, it is important to note that in addition to the scale reliabilities reported above the validity of the CKTR scales has been examined in a number of studies. Results indicate that

individual items measure what they are designed to measure and that the scale measures in each of the domains can differentiate the knowledge of experienced teachers from comparable adults who do not teach (Phelps, 2006a, 2006b).

Demographic data were collected from elementary school teacher participants of the CPDI in 2001-2002 at three time points: pre-institute; post-institute; and, follow-up. The pre- and post-institute data was gathered at the start and end of the summer CPDI institutes. Follow-up data was collected in spring of the school year following the summer institutes. Table 2 presents teacher characteristics, training, experience, as well as pre-institute knowledge levels. The distribution of gender is consistent with the distribution of gender for elementary school teachers in California as a whole. The distribution of race/ethnicity is also displayed in Table 2. Although Hispanic teachers are over-represented in the sample when compared to the current state distribution of teachers, the distribution respondents is representative of new teachers in the state. Given that we have no adequate proxy for teacher SES, we need to consider that race/ethnicity effects are likely confounded with SES effects. Table 2 also provides frequencies for teachers responding to the training and experience questions.

The last four columns of Table 2 present the pre-institute knowledge scores in logits.[7] There was considerable pre-institute variation in knowledge associated with teacher background, experience and training. These descriptive results present a static picture of pre-institute knowledge levels and demonstrate to some extent that there exist substantive differences among teacher knowledge upon entering the CPDI. However, our aim is to focus on changes in knowledge associated with CPDI participation. The teacher characteristics presented in Table 2 are used to examine potential moderating effects on CPDI effectiveness.

[7] See Phelps (2003), for a description of the methods and procedures for generating test scores.

Table 2

Teacher Characteristics and Corresponding Knowledge Scores

Teacher characteristics	Frequency	Percent	Mean Pre- Institute Knowledge level			
			CMP/KCT	CMP/CK	WA/CK	SPAN
Gender						
Male	186	10.7	-0.13	-0.16	-0.46	-0.24
Female	1,548	89.3	0.06	0.00	-0.26	-0.19
Race/Ethnicity						
White	863	49.7	0.07	0.08	-0.17	-0.27
African American	44	2.5	0.20	-0.13	-0.33	-0.21
Hispanic	298	17.2	-0.11	-0.27	-0.53	-0.11
Asian	63	3.6	0.04	-0.10	-0.29	-0.27
Other	97	5.6	-0.08	-0.05	-0.36	-0.16
Decline	372	21.4	0.07	0.00	-0.33	-0.10
Credentialing Institution						
CSU	631	54.4	-0.01	-0.02	-0.24	-0.23
Private CA	143	12.3	0.01	0.02	-0.39	-0.23
UC	129	11.1	0.22	0.12	-0.18	-0.20
Private non-CA	257	22.2	0.06	0.03	-0.26	-0.21
Credential type						
Emergency-Waiver	54	5.1	-0.21	-0.29	-0.57	-0.28
Preliminary-Intern	210	19.8	0.01	0.02	-0.39	-0.23
Full-Clear	795	75.1	0.22	0.12	-0.18	-0.20
Teach ELD in Fall						
ELD no	262	18.8	0.00	0.06	-0.28	-0.33
ELD yes	1,131	81.2	0.02	-0.03	-0.26	-0.18
Experience						
New	178	11.2	0.18	-0.02	-0.34	-0.15
1-2 yrs	226	14.3	-0.09	-0.10	-0.38	-0.22
3 or more yrs	1,179	74.5	0.03	0.01	-0.24	-0.21
Total N	1,927					

The means presented in Table 2 indicate that females make up the majority of test-takers (which is consistent with the distribution of elementary school teachers). Female teachers demonstrate higher pre-institute knowledge on all four constructs. There are also differences in pre-institute knowledge by race/ethnicity. Hispanic teachers tend to have lower pre-institute knowledge on CMP/KCT, CMP/CK, and WA/CK, but higher pre-institute knowledge on Spanish. Whites and Asians had the lowest pre-institute scores on Spanish. There were differences in pre-institute by credentialing institution as well. Teachers receiving credentials from a University of California school tend to demonstrate higher pre-institute knowledge. Teachers receiving credentials from a California State University tend to score lower than average.

Consistent with expectations, teachers with emergency or waiver credentials tend to score below average on pre-institute knowledge, while teachers with full-clear credentials tend to score highest. Pre-institute knowledge results disaggregated by teaching experience exhibit a complex pattern. In general, teachers with 1-2 years of experience tend to score lowest. New teachers demonstrate greater pre-institute knowledge in Spanish, likely due to an increased attention in recent years on teaching English language learners. The descriptive results present a static picture of pre-institute knowledge levels and demonstrate to some extent that there exist substantive differences among teacher knowledge upon entering the CPDI. Our main purpose, however, is to focus on changes in knowledge associated with CPDI participation. We use the teacher characteristics presented in Table 2 to examine potential moderating effects on CPDI effectiveness.

Analysis Methods

We seek to examine three primary questions using our teacher knowledge scores as an outcome. Is participation in the CPDIs associated with changes in teacher knowledge? Do CPDIs close existing pre-institute knowledge gaps? Are changes in teacher knowledge sustainable? In order to make inferences regarding CPDI effects on teacher knowledge, we examine teacher growth trajectories. In growth models, the participants in the study essentially serve as their own controls (Campbell & Stanley, 1963). We posit that in the absence of CPDI, teacher knowledge would have remained constant. That is, we assert that in the absence of the CPDI, all else being equal, teachers would not demonstrate significant changes in knowledge, over the period under study. This assumption is important in linking knowledge changes to the CPDI, compared to teachers not attending CPDIs, but is not critical in assessing knowledge changes among

participants over time. In order to account for differences in teacher characteristics that may moderate the effects of the intervention, covariates are used in the analyses (Pedhazur, 1982). We utilize the teacher characteristics presented in Table 2 as covariates to "adjust" initial knowledge status and knowledge growth trajectories.

Previous studies of achievement growth use covariate adjustment with a pre- and post-design to determine the effects of educational/programmatic interventions; however, two time points fail to fully capture the processes through which change takes place and are problematic for studying growth (Raudenbush & Bryk, 1987; Rogosa, Brandt, & Zimowski, 1982). This is due to (potential) measurement error in the pretest, the covariance structure of random errors, and the indirect method through which traditional analyses determined whether there are effects over time. We utilize a multilevel linear (MLM), or random effects, model where observations are nested within individuals, but time intervals need not be constant nor the same across individuals as in traditional repeated measures analyses (Raudenbush & Bryk, 2002), and the number of observations per person may vary. This latter flexibility is useful because not all of the 1,927 teachers in the analysis have three assessment occasions, but we are still able to include all of them in the analysis.[8] MLM allows flexible specification of the covariance structure at every level of the analysis (Snijders & Bosker, 1999).

The analysis is based on a three-level model where at Level 1 each teacher's development is represented by a growth trajectory that depends on a set of parameters. The outcomes are nested within teachers. At Level 2, these individual growth parameters become outcomes that depend upon teacher-level characteristics, and at Level 3, teacher characteristic effects become outcome dependent upon aggregate teacher and institute-level characteristics.

It is important to first consider a three-level model in this context because we are particularly interested in examining the effect of CPDI instructional methods on teacher knowledge status and growth. By using a three-level model, we are able to divide the variance in knowledge into within-institute, between- institute, and error components. This is particularly important because data containing multiple levels of aggregation can lead to errors in interpretation when multiple levels are ignored (Aitkin & Longford, 1986; Burstein, 1980). Ignoring the nested nature of the data and simply analyzing outcomes aggregated to the institute level upwardly biases results of teacher-level predictors, because within institute teacher-level variation is lost upon aggregation

[8] We address attrition in the results section.

(Freedman, Pisani, & Purves, 1978). Not only are teacher effects biased, but it also becomes unclear whether the estimated effects are due to group effects or whether the aggregated variables proxy for unrepresented teacher effect (Burstein, 1980).

Preliminary analyses revealed that there was no meaningful between-institute variation in knowledge growth; hence we simplify the analysis to a two level growth model. This implies that teacher outcomes do not vary by institute and that the CPDIs tend to be equally effective. The two-level model is constructed in the following manner. The Level 1 model is:

$$Y_{ti} = \pi_{0i} + \pi_{1i}\alpha_{ti} + \pi_{2i}\delta_{ti} + e_{ti}, \qquad (1)$$

where Yti is the outcome at time t for teacher i, α is a time parameter measured in assessment occasions (corresponding to the three survey administrations[9]), and δti is the deflection parameter indicating whether the assessment for teacher i at time t was conducted in the Fall (i.e., after the institutes were over and teachers had returned to teaching). If δ is statistically different from 0, then its exclusion would bias the average growth trajectory estimate. If δ is not statistically significant, then we conclude that the summer institute and the follow-up had constant effects on changes in teacher knowledge and there is no deflection of the knowledge growth trajectory. If δ is statistically significantly positive, we conclude that ideas presented during the institute made more sense once teachers had time to think about them, practice them in classrooms, and/or that institute follow-ups solidified knowledge acquired during the summer. If δ is statistically significantly negative then we conclude that teachers had not thought much about what they had learned, that practical experience did not reinforce concepts taught in the institutes, and/or that follow-ups did not solidify concepts. π_{0i} is the estimate of true pre-institute knowledge for teacher i and π_{1i} is the true change in knowledge per assessment occasion for teacher i. The teacher error is assume $N \sim (0,\sigma^2)$, which is tenable given there are only three occasions (Raudenbush & Bryk, 2002).

Although the scales have moderate reliabilities, the multilevel framework enables us to examine reliability estimates for both pre-institute knowledge and knowledge change to determine whether it is possible to detect between teacher factors associated with pre-institute knowledge and growth (Raudenbush & Bryk, 2002). Equally,

[9] We code α = (test occasion – 1). In this way π_{0i} can be interpreted as pre-institute knowledge.

important is the precision with which we estimate the growth parameter, as this determines whether we can detect individual growth (Singer & Willet, 2003). Since growth trajectories are assumed to vary among teachers, at level two for status[10] at time = 0:

$$\pi 0i = \beta 00 + \beta 01X1i + \ldots + \beta 0PXPi + r0i , \quad (2)$$

where there are p = 1 to P teacher-level predictors. For the growth trajectories[11]:

$$\pi 1i = \beta 10 + \beta 11X1i + \ldots + \beta 1PXPi + r1i , \quad (3)$$

and for the time varying covariate:

$$\pi 2i = \beta 20j + \beta 21jX2ij + \ldots + \beta 2PjXPij + r2ij , \quad (4)$$

In other words, an average initial status and an average growth trajectory are estimated. These estimates are allowed to vary among teachers, and this variation among teachers is modeled by various teacher-level predictors (e.g., gender or teaching experience).

In general, we employ the following steps in building a parsimonious model that explains teacher knowledge growth. Equations 1 through 4 are combined to build a two-level growth model, which describes knowledge growth trajectories for each teacher. The first step is to use an unconditional model (a model with only a growth parameter, but no other predictors) to examine various growth trajectories and provide baseline statistics to evaluate various Level 2 models. This also provides an estimate of the mean intercept and an estimate of the mean growth trajectory. Additionally, the unconditional model determines whether these estimates are significant and whether they vary significantly between teachers. Further, this provides an estimate of the true correlation between the initial status and the growth rate. Normal pre/post designs generally provide spurious negative correlations because the error variance of the pretest is negatively correlated with growth (Bloomquist, 1977). The next stage in the analysis is to expand the unconditional model using the available teacher information.

[10] Status is pre-institute knowledge.
[11] Growth trajectories refers to the change in knowledge from pre to post institute

Results

The descriptive results presented above indicate that there are differences among teachers' knowledge on each of the four constructs prior to attending the institutes. In order to test whether the institutes had any effect on teacher knowledge we test two growth models. The first is the unconditional model. We hypothesize that if CPDI had an effect on teacher knowledge in the four areas tested, we would expect to see statistically significant estimate for the growth term (π_1 in Equation 1). Table three presents the results of the unconditional model for each of four knowledge constructs. As noted previously, the MLM approach allows for the inclusion of all teachers with at least one assessment. This reduces, to some extent, potential bias related to attrition because all teacher scores provide information for estimating initial status and growth. Of the 1,927 teachers in the sample, 351 completed all three assessments, 938 completed two of three (with 102 of those completing the second two assessments). The remaining teachers completed one assessment. Preliminary analyses indicate that the average pre-institute knowledge score is not related to the number of assessments a teacher completed. Further, separate analysis by number of assessments completed reveals that the overall patterns we report below are consistent irrespective of the number of assessments completed. The robustness of results provides some evidence that teacher attrition does not affect inferences.

Table 3

Baseline Estimates of Teacher Knowledge and Knowledge Growth (In Logits)

	CMP/KCT		CMP/CK		WA/CK		SPAN	
Fixed effects	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
Pre-Institute Knowledge level	0.03	0.02	-0.02	0.02	-0.28	0.02 *	-0.20	0.01 *
Average growth in Knowledge	0.08	0.02 *	0.08	0.01 *	0.17	0.01 *	0.01	0.01
Post-Institute Knowledge change	-0.14	0.02 *	-0.12	0.02 *	-0.20	0.02 *	0.02	0.01
Random effects	Variance	x^2	Variance	x^2	Variance	x^2	Variance	x^2
Pre-Institute Knowledge level	0.37	8,966 *	0.40	10,070 *	0.45	11,543 *	0.13	5,343
Average growth in Knowledge	0.26	4,298 *	0.12	3,104 *	0.18	3,825 *	0.00	1,890 *
Post-Institute Knowledge change	0.58	3,672 *	0.27	2,805 *	0.41	3,293 *	0.01	1,901 *

Due to the scaling metric, we are less interested in the actual estimates for pre-institute teacher knowledge and the corresponding statistical significance, as we are for the associated variance components (random effects). The random effect estimates indicate that there were significant differences in the levels of teacher knowledge upon entering the institutes in three of the knowledge constructs. Teacher knowledge in Spanish did not vary significantly among teachers. The results for the average growth estimates indicate that teachers demonstrate statistically significant knowledge growth in three of the four knowledge constructs. There was no significant growth in teacher knowledge of Spanish.[12] The variance components associated with the growth estimates indicate that for all four constructs growth varied significantly among teachers. That is, teachers exhibited significantly varied responses to CPDI training as measured varying growth rates.

We are particularly interested in post-institute knowledge changes (as represented by δ). Again except for Spanish, the post institute effect on growth was statistically significantly negative ($\delta < 0$). This indicates that teachers had lower levels of knowledge on CMP/KCT, CMP/CK, and WA/CK at the time of the follow-up than they at the end of CPDI instruction. In other words, they appeared to have forgotten some of what they learned and that practical classroom experience and follow-ups did not solidify knowledge acquired during the summer. Had we not included this parameter in the

[12] Our inability to detect growth in S may certainly be due to the fact that this scale had the only moderate reliability (the lowest of the four scales).

model, it would have biased the growth estimate (in this case) down, and it would have appeared as if teachers had gained less from CPDI instruction than they actually did. The associated variance components indicate that this knowledge drop varied significantly among teachers, as well.

Table 4 presents the substantive effects of the variance components on initial status, growth and post summer knowledge change. Table 4 also presents quasi-effect size estimates. The effect size estimates for initial status compare teachers who are one standard deviation above average in pre-institute knowledge to teachers who are one standard deviation below average in pre-institute knowledge.[13] The effect size estimates for growth use the temporal variation in growth as the denominator (Raudenbush & Xiao-Feng, 2001). In this case the effect size estimates describe the difference between the relative difference in growth between teachers who are one s.d. above average and teachers who are one s.d. below average in growth.

Table 4

Differences in Pre-Institute Knowledge and Growth from Baseline Estimates

	CMP/KCT Estimate	Effect Size	CMP/CK Estimate	Effect Size	WA/CK Estimate	Effect Size	SPAN Estimate	Effect Size
Pre-Institute Knowledge level								
- I s.d.	-0.58		-0.65		-0.95		-0.56	
+ I s.d.	0.64	1.74	0.61	1.78	0.39	1.80	0.16	1.65
Average growth in Knowledge								
- I s.d.	-0.42		-0.27		-0.26		-0.02	
+ I s.d.	0.59	3.19	0.42	2.26	0.59	2.86	0.05	0.25
Post-Institute Knowledge change								
- I s.d.	-0.91		-0.65		-0.84		-0.06	
+ I s.d.	0.62	4.80	0.40	3.45	0.44	4.28	0.11	0.66

The results in Table 4 indicate that there is a substantively large amount of variation in growth and post institute knowledge change for all but the Spanish. knowledge construct. In fact, for CMP/KCT, CMP/CK, and WA/CK, there is relatively

[13] This is similar to a standard effect size estimate that compares the difference in the outcome between two groups divided by the s.d of the control group – yielding the s.d. difference between the groups (Cooper & Hedges, 1994).

more variation among teachers in knowledge change than there is in knowledge levels upon entering the CPDIs. Further, post-institute knowledge changes demonstrated the largest of relative variation. In other words, teachers entered the CPDI with varying amounts of knowledge, benefited from CPDI at varying rates, and then post-institute, they retained knowledge at even greater varying rates.

The next step in the analysis attempts to discern which teacher characteristics, experience, and training variables might be associated with pre-institute knowledge levels and knowledge change. These results are presented in Tables 4, 5, 6, and 7 for CMP/KCT, CMP/CK, WA/CK, and Spanish, respectively. Each knowledge domain was initially modeled with the same set of variables, but we present only the most parsimonious models in the tables.

Table 5A

Estimated Effects of Teacher Covariates on Pre-Institute CMP/KCT Knowledge

Fixed effects	Coeff.	SE	p-value
Mean Pre-Institute	0.03	0.02	0.06
Difference between:			
Female and Male	0.17	0.05	0.00
Asian and White	-0.07	0.08	0.41
Decline and White	-0.06	0.05	0.23
Other and White	-0.18	0.07	0.02
Hispanic and White	-0.18	0.04	0.00
Black and White	0.09	0.08	0.24
Cred. Ins. Private CA and CSU	0.04	0.06	0.54
Cred. Ins. UC and CSU	0.26	0.06	0.00
Cred. Ins. Out of State and CSU	0.06	0.04	0.16
Emergency-Waiver and Full-Clear			
Preliminary-Intern and Full-Clear			
Teach ELL Fall and Not Teach ELL Fall			
1-2 Yrs Teaching Exp. And 0 Yrs Teaching	-0.28	0.07	0.00
3+ Yrs Teaching Exp. And 0 Yrs Teaching	-0.17	0.06	0.00

Table 5A presents the results for pre-institute CMP/KCT knowledge. The results indicate that initial pre-institute knowledge levels were moderated by several teacher characteristics. Males had higher CMP/KCT knowledge levels than females and whites tended to have higher CMP/KCT knowledge levels than non-whites. Both experience and training variables moderated the pre-institute knowledge levels. In fact, the two most important determinants of pre-institute knowledge were where a teacher was credentialed and how much teaching experience they had. Teachers trained at a UC (University of California) scored about 0.26 logits higher than teachers trained at a CSU (California State University), ceteris paribus. This equates to an effect size of about 0.36. Another important determinant of pre-institute CMP/KCT knowledge levels is teaching experience; although the effect is, perhaps, contrary to expectations. New teachers demonstrated greater pre-institute CMP/KCT knowledge than more experienced teachers. This may be due to the fact that new teachers are closer to "book" knowledge, as they recently completed their teacher education program. Alternatively, this may simply be a cohort effect in that the knowledge items are more closely aligned with current teacher training curriculum.

Table 5B presents results for changes CMP/KCT knowledge. Several teacher characteristic and experience indicators are related to knowledge growth. The mean knowledge growth is 0.08 logits per test occasion. This means that, on average, teachers would gain about 0.08 logits, or about 0.11 s.d.s in knowledge per assessment. Males gained significantly less than females and whites tended to gain more than non-whites. Teachers who were not scheduled to teach ELL students in the fall gained less than teachers who were scheduled to teach ELL students in the fall. Teachers with either 1-2 years of teaching experience or 3 or more years of experience demonstrated significantly greater CMP/KCT knowledge gains than new teachers.

Table 5B also presents the results for the post-institute effect on knowledge. In other words, after returning to teaching, how much did teachers retain from the CPDI training? As noted above, we defined the post-institutes as being neutral, additive, or negative. The results in Table 5 indicate that knowledge changes post institute varied by teacher characteristics and experience. Post-institute knowledge was additive for Hispanic and African American and Asian teachers. Teachers indicating that they would be teaching ELL students in the fall also experienced a post-institute additive effect in CMP/KCT knowledge change. Teachers with no experience had a negative effect, while teachers with 1-2 years and 3 or more years of experience had additive effects.

Table 5B

Estimated Effects of Teacher Covariates on Changes in CMP/KCT Knowledge

Fixed effects	Coeff.	SE	p-value
Mean Change in Knowledge	0.08	0.02	0.00
Difference between:			
Female and Male	-0.11	0.05	0.03
Asian and White	-0.15	0.09	0.09
Decline and White	-0.11	0.04	0.02
Other and White	0.03	0.07	0.68
Hispanic and White	-0.12	0.04	0.01
Black and White	-0.20	0.07	0.00
Cred. Ins. Private CA and CSU			
Cred. Ins. UC and CSU			
Cred. Ins. Out of State and CSU			
Emergency-Waiver and Full-Clear			
Preliminary-Intern and Full-Clear			
Teach ELL Fall and Not Teach ELL Fall	-0.16	0.05	0.00
1-2 Yrs Teaching Exp. And 0 Yrs Teaching	0.32	0.07	0.00
3+ Yrs Teaching Exp. And 0 Yrs Teaching	0.29	0.06	0.00
Mean Difference in Post-institute Knowledge	-0.14	0.02	0.00
Difference between:			
Female and Male	0.15	0.08	0.07
Asian and White	0.23	0.13	0.09
Decline and White	0.18	0.07	0.01
Other and White	-0.01	0.12	0.91
Hispanic and White	0.20	0.07	0.01
Black and White	0.33	0.12	0.01
Cred. Ins. Private CA and CSU			
Cred. Ins. UC and CSU			
Cred. Ins. Out of State and CSU			
Emergency-Waiver and Full-Clear			
Preliminary-Intern and Full-Clear			
Teach ELL Fall and Not Teach ELL Fall	0.27	0.08	0.00
1-2 Yrs Teaching Exp. And 0 Yrs Teaching	-0.47	0.11	0.00
3+ Yrs Teaching Exp. And 0 Yrs Teaching	-0.41	0.09	0.00

Table 6A presents the results for pre-institute CMP/CK knowledge. Again, only the most parsimonious model is presented. Teacher characteristics, training, and experience tended to be less associated with CMP/CK knowledge and CMP/CK knowledge changes.

Table 6A

Estimated Effects of Teacher Covariates on Pre-Institute CMP/CK Knowledge

Fixed effects	Coeff.	SE	p-value
Mean Pre-Institute	-0.02	0.02	0.16
Difference between:			
Female and Male	0.11	0.06	0.04
Asian and White	-0.19	0.09	0.04
Decline and White	-0.07	0.04	0.11
Other and White	-0.12	0.07	0.11
Hispanic and White	-0.33	0.05	0.00
Black and White	-0.18	0.08	0.02
Cred. Ins. Private CA and CSU	0.06	0.07	0.37
Cred. Ins. UC and CSU	0.19	0.06	0.00
Cred. Ins. Out of State and CSU	0.05	0.04	0.22
Emergency-Waiver and Full-Clear	-0.24	0.08	0.00
Preliminary-Intern and Full-Clear	-0.06	0.05	0.17
Teach ELL Fall and Not Teach ELL Fall			
1-2 Yrs Teaching Exp. And 0 Yrs Teaching			
3+ Yrs Teaching Exp. And 0 Yrs Teaching			

There are again significant differences in pre-institute knowledge in CMP/CK as there were for CMP/KCT. Teachers pre-institute knowledge varied by both gender and race/ethnicity. Teachers receiving their teacher training at a UC scored about 0.27 s.d.s higher than teachers trained at a CSU, ceteris paribus. Further, teachers with preliminary or credential waiver scored about 0.34 s.d.s lower than teachers with full/clear credentials. Unlike for CMP/KCT, teacher experience was unrelated to pre-institute CMP/CK.

Table 6B presents results for changes in CMP/CK knowledge. Gender and race/ethnicity was unrelated to knowledge growth. Teachers who received their

training at a UC demonstrated slightly greater CMP/CK knowledge growth than teachers trained at a CSU. Among the teacher characteristics, training and experience variables, only two related to post-institute CMP/CK knowledge change. African Americans demonstrated an additive effect as their CMP/CK knowledge benefited from the additional post-institute experience.

Table 6B

Estimated Effects of Teacher Covariates on Changes in CMP/CK Knowledge

Fixed effects	Coeff.	SE	p-value
Mean Change in Knowledge	0.08	0.01	0.00
Difference between:			
Female and Male			
Asian and White	0.07	0.06	0.22
Decline and White	-0.10	0.03	0.01
Other and White	-0.09	0.05	0.08
Hispanic and White	-0.05	0.04	0.15
Black and White	-0.08	0.04	0.08
Cred. Ins. Private CA and CSU	-0.01	0.02	0.71
Cred. Ins. UC and CSU	0.08	0.02	0.00
Cred. Ins. Out of State and CSU	0.01	0.01	0.42
Emergency-Waiver and Full-Clear			
Preliminary-Intern and Full-Clear			
Teach ELL Fall and Not Teach ELL Fall			
1-2 Yrs Teaching Exp. And 0 Yrs Teaching			
3+ Yrs Teaching Exp. And 0 Yrs Teaching			
Mean Difference in Post-institute Knowledge	-0.12	0.02	0.00
Difference between:			
Female and Male			
Asian and White	-0.15	0.09	0.11
Decline and White	0.14	0.06	0.02
Other and White	0.13	0.08	0.12
Hispanic and White	0.04	0.06	0.48
Black and White	0.18	0.07	0.01
Cred. Ins. Private CA and CSU			
Cred. Ins. UC and CSU			
Cred. Ins. Out of State and CSU			
Emergency-Waiver and Full-Clear			
Preliminary-Intern and Full-Clear			
Teach ELL Fall and Not Teach ELL Fall			
1-2 Yrs Teaching Exp. And 0 Yrs Teaching			
3+ Yrs Teaching Exp. And 0 Yrs Teaching			

In general, however, while the CPDI engendered knowledge growth in CMP/CK, it did not facilitate teachers closing knowledge gaps, where they existed. Further, the mean overall gain was only about 0.06 s.d.s.

Table 7A presents the results of the model describing pre-institute WA/CK knowledge. Again we examine the full set of teacher characteristic, experience, and training variables, but present only the final parsimonious model. The results in Table 7A indicate that pre-institute WA/CK knowledge was related to several teacher and training variables. There were significant difference by gender and race/ethnicity. Male teachers scored about 0.29 s.d.s higher than female teachers. White teachers demonstrated greater pre-institute WA/CK knowledge than all but Asian teachers. The greatest discrepancy was between Whites and Hispanics (0.49 s.d. difference). There was no difference between teachers receiving training at a UC or CSU; however, teachers receiving training at a private CA university scored significantly lower than teachers receiving training at a CSU. As with the previous knowledge constructs, teachers with emergency or waiver credentials began the institutes significantly behind (0.29 s.d.s) teachers with full/clear credentials.

Table 7A

Estimated Effects of Teacher Covariates on Pre-Institute WA/CK Knowledge

Fixed effects	Coeff.	SE	p-value
Mean Pre-Institute	-0.28	0.02	0.00
Difference between:			
Female and Male	0.20	0.06	0.00
Asian and White	-0.13	0.08	0.14
Decline and White	-0.16	0.05	0.00
Other and White	-0.16	0.07	0.02
Hispanic and White	-0.34	0.05	0.00
Black and White	-0.13	0.07	0.08
Cred. Ins. Private CA and CSU	-0.15	0.06	0.02
Cred. Ins. UC and CSU	0.08	0.07	0.26
Cred. Ins. Out of State and CSU	-0.02	0.05	0.72
Emergency-Waiver and Full-Clear	-0.20	0.09	0.03
Preliminary-Intern and Full-Clear	-0.04	0.06	0.47
Teach ELL Fall and Not Teach ELL Fall			
1-2 Yrs Teaching Exp. And 0 Yrs Teaching			
3+ Yrs Teaching Exp. And 0 Yrs Teaching			

Table 7B presents results for changes in WA/CK knowledge. Growth in WA/CK knowledge was relatively higher for WA/CK than for CMP/KCT, or CMP/CK. There was little differentiation in growth among the teacher variables, except that White teachers tended to demonstrate faster achievement growth than all but Asian and other race/ethnicity groups.

The post institute knowledge change for WA/CK was also relatively greater (negative) than for either CMP/KCT or CMP/CK. Race/ethnicity and teacher training were related to the post-institute knowledge change. African Americans demonstrated an additive effect, which reduced the gap with white teachers, although not completely. Teachers trained at a UC experienced a post-institute additive effect as well. This effect created a 0.34 s.d gap between teachers receiving credentials from a UC and teachers receiving credentials at a CSU.

Table 8A presents the results for pre-institute Spanish knowledge. The results for Spanish are considerably different than for the other three knowledge domains that

were tested. While Spanish knowledge was related to various teacher training, experience, and training variables, teachers demonstrated no change in Spanish knowledge. The results in Table 8A indicate that Hispanics had greater Spanish knowledge than whites, pre-institute. Further, teachers who planned to teach EL students following the institute also had greater pre-institute knowledge in Spanish. Table 8B presents results for changes in Spanish knowledge. As noted, teachers demonstrated no growth in Spanish knowledge and this growth was relatively unaffected by teacher characteristics. Teachers who planned to teach ELL students demonstrated less growth than teachers who did not plan to teach ELL students. Also, teachers with 3 or more years of experience demonstrated more growth than first year teachers.

Table 7B

Estimated Effects of Teacher Covariates on Changes in WA/CK Knowledge

	Coeff.	SE	p-value
Mean Change in Knowledge	0.17	0.01	0.00
Difference between:			
Female and Male			
Asian and White	-0.08	0.07	0.28
Decline and White	-0.07	0.04	0.05
Other and White	-0.12	0.06	0.06
Hispanic and White	-0.09	0.04	0.02
Black and White	-0.18	0.04	0.00
Cred. Ins. Private CA and CSU	0.10	0.06	0.09
Cred. Ins. UC and CSU	-0.11	0.06	0.06
Cred. Ins. Out of State and CSU	-0.03	0.04	0.44
Emergency-Waiver and Full-Clear			
Preliminary-Intern and Full-Clear			
Teach ELL Fall and Not Teach ELL Fall			
1-2 Yrs Teaching Exp. And 0 Yrs Teaching			
3+ Yrs Teaching Exp. And 0 Yrs Teaching			
Mean Difference in Post-institute Knowledge	-0.20	0.02	0.00
Difference between:			
Female and Male			
Asian and White	0.04	0.12	0.74
Decline and White	0.12	0.06	0.04
Other and White	0.16	0.11	0.13
Hispanic and White	0.13	0.06	0.05
Black and White	0.26	0.08	0.00
Cred. Ins. Private CA and CSU	-0.10	0.09	0.28
Cred. Ins. UC and CSU	0.24	0.10	0.02
Cred. Ins. Out of State and CSU	0.04	0.07	0.59
Emergency-Waiver and Full-Clear			
Preliminary-Intern and Full-Clear			
Teach ELL Fall and Not Teach ELL Fall			
1-2 Yrs Teaching Exp. And 0 Yrs Teaching			
3+ Yrs Teaching Exp. And 0 Yrs Teaching			

As there was no knowledge growth, on average, there was also no post-institute knowledge effect. However, teachers who were teaching ELLs demonstrated an additive effect, and new teachers demonstrated a significantly lower negative effect than teachers with 1 or more years of experience.

Table 8A

Estimated Effects of Teacher Covariates on Pre-Insatiate Spanish Knowledge

Fixed effects	Coeff.	SE	p-value
Mean Pre-Institute	-0.20	0.01	0.00
Difference between:			
Female and Male			
Asian and White	0.00	0.06	0.95
Decline and White	0.17	0.03	0.00
Other and White	0.10	0.05	0.07
Hispanic and White	0.14	0.02	0.00
Black and White	0.06	0.05	0.24
Cred. Ins. Private CA and CSU			
Cred. Ins. UC and CSU			
Cred. Ins. Out of State and CSU			
Emergency-Waiver and Full-Clear			
Preliminary-Intern and Full-Clear			
Teach ELL Fall and Not Teach ELL Fall	0.13	0.03	0.00
1-2 Yrs Teaching Exp. And 0 Yrs Teaching			
3+ Yrs Teaching Exp. And 0 Yrs Teaching			

Table 8B

Estimated Effects of Teacher Covariates on Changes in SPAN Knowledge

Fixed effects	Coeff.	SE	p-value
Mean Change in Knowledge	0.02	0.01	0.07
Difference between:			
Female and Male			
Asian and White	-0.07	0.05	0.13
Decline and White	-0.10	0.02	0.00
Other and White	0.02	0.04	0.58
Hispanic and White	-0.04	0.02	0.07
Black and White	-0.07	0.04	0.13
Cred. Ins. Private CA and CSU			
Cred. Ins. UC and CSU			
Cred. Ins. Out of State and CSU			
Emergency-Waiver and Full-Clear			
Preliminary-Intern and Full-Clear			
Teach ELL Fall and Not Teach ELL Fall	-0.06	0.03	0.03
1-2 Yrs Teaching Exp. And 0 Yrs Teaching	0.05	0.03	0.09
3+ Yrs Teaching Exp. And 0 Yrs Teaching	0.07	0.02	0.01
Mean Difference in Post-institute Knowledge	0.02	0.01	0.12
Difference between:			
Female and Male			
Asian and White	0.08	0.09	0.40
Decline and White	0.19	0.04	0.00
Other and White	0.00	0.07	0.95
Hispanic and White	0.06	0.04	0.10
Black and White	0.04	0.05	0.46
Cred. Ins. Private CA and CSU			
Cred. Ins. UC and CSU			
Cred. Ins. Out of State and CSU			
Emergency-Waiver and Full-Clear			
Preliminary-Intern and Full-Clear			
Teach ELL Fall and Not Teach ELL Fall	0.12	0.05	0.01
1-2 Yrs Teaching Exp. And 0 Yrs Teaching	-0.10	0.05	0.05
3+ Yrs Teaching Exp. And 0 Yrs Teaching	-0.11	0.04	0.01

Conclusion and Discussion

In this analysis we examined whether the CPDIs affected changes in teacher knowledge in four specific domains: Comprehension content knowledge (CMP/CK); Comprehension knowledge of content and teaching (CMP/KCT); Word analysis content knowledge (WA/CK); and, Spanish language (SPAN). Each of the four knowledge domains was not emphasized equally during the institutes. The institutes focused primarily on CMP/CK, although both CMP/KCT and WA/CK were included in the training as well. Spanish received the least amount of coverage and teachers also felt that the training in Spanish was not as strong as it could have been (Griffin et al., 2003b).

We utilized longitudinal growth models and hypothesized that statistically significant estimates for the growth parameters would indicate that teacher knowledge was affected by CPDI attendance. Teachers exhibited statistically significant knowledge growth in three of the four constructs (CMP/KCT, CMP/CK, and WA/CK). Only Spanish knowledge was unaffected and this is consistent with teacher evaluation of the institutes in this area and may also reflect the less reliable equating and scale measure in this domain.

Taking advantage of the longitudinal nature of the data, we further divided the changes in knowledge into growth exhibited during the week of training, and the period between when the institute ended and the post-institute follow-up occurred. This allowed us to examine whether teachers continued to gain content knowledge through practical experience applied to lessons learned during the institutes, or whether teachers would demonstrate difficulty in retaining content knowledge. In general, the results demonstrated that teachers tended to recall less of the content at the post-institute administration than they had at the end-of-institute administration.

We further examined correlates of knowledge change by including various teacher demographic, training, and experience variables in the models. We also examined the relationship between teacher CPDI evaluation, and teacher reported practices and changes in knowledge. In general, there was significant variation in teacher knowledge upon entering the CPDIs. Teacher knowledge varied by 1.7 standard deviations (+/- one standard deviation) pre-institute between 0.7 to 4.8 standard deviations post institute. The pattern of this variation among the knowledge domains was relatively similar except for Spanish, which demonstrated less variation in both pre-institute variation as well in knowledge change.

Pre-institute knowledge differences were related, in varying degrees, to teacher demographic, training, and experience variables. Accounting for differences in teacher training and experience, there existed significant gaps in pre-institute knowledge by race/ethnicity. Given that the models tested/explained only a small fraction of the pre-institute variance in teacher knowledge, research is warranted to ascertain which teacher characteristics, not related to the experience and training variables already included, might account for the pre-institute gaps in knowledge. Also, the pattern of pre-institute knowledge differences was not uniform among the knowledge domains. While statistical significance varied among the domains, the estimated coefficients were consistent in terms of direction.

Taken together, we make several inferences regarding teacher training and experience. That is, that there are substantive differences among teachers. In general, the credentialing institution is associated with differences in pre-institute knowledge. We find that teachers trained at UC schools of education tend to have higher pre-institute knowledge than their colleagues trained at CSUs. Selection and opportunities related to family SES may affect which credentialing institution a teacher attended; however, our focus is not on whether credentialing institutions caused existing pre-institute knowledge gaps—rather than upon exiting those institutions, gaps exist. We infer that the type of credential matters, as teachers with emergency/waiver credentials tend to score lower than teachers with full/clear credentials. Interestingly, teachers with preliminary/intern credentials are indistinguishable from teachers with full/clear credentials. Teaching experiences plays a role as well. Based on the results we posit that new teachers, who are less removed from formal teacher education, fare better on the pre-institute test, but that teachers with at least one year of experience benefit more from the CPDI. Teachers also demonstrate that they self-select important information; that is, teachers who will be teaching ELL students demonstrate greater pre-institute knowledge in Spanish. The initial differences in knowledge are interesting in that they point to potential differences that may exist among teachers in general. Whether these specific patterns are generalizable to the entire population of California teachers is perhaps less important than the fact that there exists substantial variation in knowledge. Given the common knowledge assertion that teachers matter in student success, this variation in knowledge could go a long way towards explaining between teacher variation in student academic achievement. Additional research linking teachers to students and examining the relationship between teacher knowledge and student academic achievement is certainly warranted.

As noted, attending the CPDI resulted in statistically significant knowledge changes in three of the four tested domains. Further, these changes were broken out between growth during the institute and after the institute. The results do not provide one over-arching pattern, rather relationships between covariates and knowledge (initial status and change) tended to be more similar within a construct than among them. The pattern of significant predictors for pre-institute, growth, and post-institute change, were generally similar within a domain, rather than any single predictor being significant across the knowledge domains. The results indicate that teachers with different characteristics demonstrated different growth patterns. This was most noticeably evident in the effect of practical experience on post-institute knowledge change. As expected, some teachers had an additive effect—that is performed better on the knowledge tests at the follow-up—while others had a negative effect—that is, failed to retain all of the knowledge they had acquired.

We also examined the extent to which teacher perceptions and practices were related to knowledge. Both pre-institute knowledge and knowledge changes tended to be invariant to either teacher evaluations or teacher reported practices. From the standpoint of the CPDI, this could be considered a positive sign in that the institutes had a uniform effect on teachers across these characteristics.

However, from another perspective this uniform effect is a matter of some concern. This implies that teacher perceptions and evaluations of institute quality are poor indicators of professional development quality and this unreliability further reduces the likelihood of finding systematic results between in-service professional development and student outcomes. Further, the institutes were designed in large part to ensure that all students receive high-quality instruction and can learn to read by the end of the third grade. This is particularly a concern for disadvantaged students who traditionally are taught by the least qualified teachers (Darling-Hammond, 2004). If the CPDIs succeeded in this goal the gap between less and more knowledgeable teachers would narrow and not grow. Indeed, the results suggest that the teachers who benefited most from the CPDI training were those who already knew the most.

Overall, the CPDIs are related to changes in teacher knowledge; however, some of these gains were lost over time. The analyses we presented only explain a small proportion of the variation in teacher knowledge. This implies that further research is warranted in order to determine which other factors may affect how much teachers benefit from attending institutes, both in the short term, and in the long run. Further

research should examine post-institute support mechanisms that might reduce the lack of knowledge retention.

It is encouraging that teachers attending the CPDI institutes showed growth in their knowledge. However, much work is still needed to understand teacher knowledge itself and its relation to effective teaching practice and ultimately student achievement. The knowledge measures used in this study are promising since they are sensitive to both differences among teachers and to professional development. At the same time, it is important to recognize that teacher performance on these measures ultimately only matters if it is associated with valued educational outcomes such as improvement in instruction and student achievement. An important next step for research of the type conducted here is to better understand the teacher knowledge measures themselves. This includes both validation against outcomes such as student achievement and related research to understand whether the CPDIs affected significant growth that went unmeasured. Progress on these important questions will allow researchers to not only ask how much and for how long professional development affects teacher knowledge, but also how the extent to which these changes in teacher quality matter for improving the quality of instruction.

References

Aitkin, M., & Longford, N. (1986). Statistical modeling issues in school effectiveness studies. *Journal of the Royal Statistical Society, 149*(1), 1-43.

Ball, D. L. (1988). *Knowledge and reasoning in mathematical pedagogy: Examining what prospective teachers bring to teacher education*. Unpublished doctoral dissertation, Michigan State University, East Lansing.

Ball, D. L. (1991). Research on teaching mathematics: Making subject matter part of the equation. In J. Brophy (Ed.), *Advances in research on teaching*, *2*, 1-48. Greenwich, CT: JAI Press.

Ball, D. L., Lubienski, S., & Mewborn, D. (2001). Research on teaching mathematics: The unsolved problem of teachers' mathematics knowledge. In V. Richardson (Ed.), *Handbook of research on teaching*, 4th edition (pp. 433-456). New York: Macmillan.

Bloomquist, N. (1977). On the relation between change and initial value. *Journal of the American Statistical Association*, *72,* 746-749.

Bos, C. S., Mather, N., & Dickson, S. (2001). Perceptions and knowledge of preservice and inservice educators about early reading instruction. *Annals of Dyslexia, 51,* 97-120.

Brady, S., & Moats, L. (1997). *Informed instruction for reading success: Foundations for teacher preparation* (Opinion paper 120). Baltimore, MD: International Dyslexia Association.

Burstein, L. (1980). The analysis of multi-level data in educational research and evaluation. *Review of Research in Education*, *4,* 158-233.

Campbell, D. T., & Stanley, J. (1963). *Experimental and quasi experimental designs for research*. Chicago: Rand McNally.

Carlisle, J., Correnti, R., Phelps, G., & Zeng, J. (2006). Investigating teachers' knowledge of language structure and its relation to students' reading achievement. *Reading and writing: An interdisciplinary journal.* Submitted for publication.

Cunningham, A. E., Perry, K. E., Stanovich, K. E., Stanovich, P. J., & Chappell, M. (2001). *Is teachers' knowledge of important declaritive knowledge of reading well calibrated?* Paper presented at the Annual Meeting of the Society for the Scientific Study of Reading, Denver, CO.

Darling-Hammond, L. (1999). *Solving the dilemmas of teacher supply, demand, and standards: How we can ensure a competent, caring, and qualified teacher for every child*. Washington, DC: National Commission on Teaching and America's Future.

Darling-Hammond, L. (2004). Inequality and the right to learn: Access to qualified teachers in California's public schools. *Teachers College Record, 106*(10), 1936-1966.

Freedman, D., Pisani, R., & Purves, R. (1978). *Statistics*. New York: WW Norton and Company.

Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal, 38*(4), 915-945.

Gess-Newsome, J., & Lederman, N. G. (1995). Biology teachers' perceptions of subject matter structure and its relationship to classroom practice. *Journal of Research in Science Teaching, 32*(3), 301-325.

Griffin, N., Aguirre-Muñoz, Z., Miyoshi, J., Roberson, I., & Amabisca, A., (2001, November). *Evaluation of the California Professional Development Institutes in English language arts: Grades pre K-6.* Center for the Study of Evaluation (CSE), University of California, Los Angeles.

Griffin, N., Aguirre-Muñoz, Z., Goldschmidt, P., Amabisca, A., Miyoshi, J., Swigert, S. et al. (2003a, February). *Evaluation of the California Professional Development Institutes in English language arts: Primary grades year two report.* Center for the Study of Evaluation (CSE), University of California, Los Angeles.

Griffin, N., Aguirre-Muñoz, Z., Goldschmidt, P., Amabisca, A., Miyoshi, J., Swigert, S. et al. (2003b, November). *Evaluation of the secondary school California Professional Development Institutes in English language arts: 2003 Addendum Report.* Center for the Study of Evaluation (CSE), University of California, Los Angeles.

Grossman, P. L. (1990). A study in contrast: Sources of pedagogical content knowledge for secondary English. *Journal of Teacher Education, 40*(5), 24-31.

Grossman, P. L. (1991). What are we talking about anyway? Subject-matter knowledge of secondary English teachers. In J. Brophy (Ed.), *Advances in research on teaching, 2*, 245-264. Greenwich, CT: JAI Press.

Hambleton, R., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. London: Sage Publications.

Leinhardt, G., & Smith, D. (1985). Expertise in mathematics instruction: Subject matter knowledge. *Journal of Educational Psychology, 77*(3), 247-271.

Magnusson, S., Krajcik, J., & Borko, H. (1999). Nature, sources and development of pedagogical content knowledge for science teaching. In N. G. Lederman (Ed.), *Examining pedagogical content knowledge: The construct and its implications for science education* (pp. 95-132). Dordrecht: Kluwer Academic Publishers.

McCutchen, D., Abbott, R. D., Green, L. B., Beretvas, S. N., Cox, S., Quiroga, T., Potter, N. S., & Gray, A. L. (2002). Beginning literacy: Links among teacher knowledge, teacher practice, and student learning. *Journal of Learning Disabilities, 35*(1), 69-87.

McCutchen, D., & Berninger, V. (1999). Those who know, teach well: Helping teachers master literacy-related subject-matter knowledge. *Learning Disabilities Research and Practice, 14*(4), 215-226.

McCutchen, D., Harry, D. R., & Cox, S. (2002). Reading teachers knowledge of children's literature and English phonology. *Annals of Dyslexia, 52,* 207-228.

Mislevy, R. J., & Bock, R. D. (1997). *Bilog: Item analysis and test scoring with binary models*. Lincolnwood, IL: Scientific Software International.

Moats, L. (1994). The missing foundation in teacher education: Knowledge of the structure of spoken and written language. *Annals of Dyslexia, 44*, 81-102.

Moats, L. (2000). *Speech to print: Language essentials for teachers*. Baltimore, MD: Paul H. Brooks Publishing Co.

Moats, L., & Lyon, R. (1996). Wanted: Teachers with knowledge of language. *Topics of Language Disorders, 16*(2), 73-86.

National Board for Professional Teaching Standards (NBPTS). (2001). *Middle childhood generalist standards,* 2nd edition. Washington, DC: Author.

National Center for Education Statistics. (1999). *Teacher quality: A report on the preparation and qualifications of public school teachers*. Washington DC: U.S. Department of Education.

No Child Left Behind Act of 2001, Pub. L. No. 107–110 (2002).

Pedhazur, E. (1982). *Multiple regression in behavioral research*. Fort Worth: Harcourt Brace Jovanovich.

Phelps, G. (2003). *CPDI 2001/2002 Measures Development*. The Study of Instructional Improvement & The Learning Mathematics for Teaching Project, University of Michigan.

Phelps, G. (2005). *Content knowledge for teaching reading.* Unpublished doctoral dissertation, University of Michigan, Ann Arbor.

Phelps, G. (2006a). Just knowing how to read isn't enough! What teachers know about the content of reading. *Reading Research Quarterly*. Submitted for publication.

Phelps, G. (2006b). *Investigating the validity of content knowledge for teaching reading measures: A consistency analysis and study packet.* Ann Arbor: University of Michigan: Consortium for Policy Research in Education, Study of Instructional Improvement, Research Report.

Phelps, G., & Schilling, S. (2004). Developing measures of content knowledge for teaching reading. *Elementary School Journal, 105*(1), 31-48.

Raudenbush, S. W., & Bryk, A. S. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin*, *104*(3), 396-404.

Raudenbush, S. W., & Bryk, A. S. (2002) *Hierarchical linear models: Applications and data analysis methods,* 2nd ed. Thousand Oaks, CA: Sage Publications.

Raudenbush, S., & Xiao-Feng, L. (2001). Effects of study duration, frequency of observation, and sample size on power in studies of group differences in polynomial change. *Psychological Methods, 6*(4), 387-401.

Rogosa, D. R., Brandt, D., & Zimowski, M. (1982). A growth curve approach to the measurement of change. *Psychological Bulletin*, *92,* 726-74.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4-14.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57*, 1-22.

Singer, J., & Willet, J. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford University Press.

Snijders, T. A. B., & Bosker, R. J. (1999) *Multilevel analysis: An introduction to basic and advanced multilevel modeling.* London: Sage Publishers.

Thissen, D., & Wainer, H. (Eds.). (2001). *Test scoring*. Mahwah, NJ: Lawrence Erlbaum Associates.

Wilson, S., Shulman, L., & Richert, A. (1987). "150 different ways of knowing": Representations of knowledge in teaching. In J. Calderhead (Ed.), *Exploring teachers' thinking* (pp. 104-123). Eastbourne, England: Cassell.

Wilson, S., & Wineburg, S. (1988). Peering at history through different lenses: The role of disciplinary perspectives in teaching history. *Teachers College Record, 89*(4), 525-539.

Wineburg, S., & Wilson, S. M. (1991). The subject matter knowledge of history teachers. In J. Brophy (Ed.), *Advances in research on teaching, 2*, 305-345. Greenwich CT: JAI Press.

Wong-Fillmore, L., & Snow, C. E. (2002). What teachers need to know about language. In D. Christian (Ed.), *What teachers need to know about language* (pp. 7-54). McHenry, IL: Delta Systems Co., Inc.