NCEE 2009-4043

U.S. DEPARTMENT OF EDUCATION

An Evaluation of Teachers Trained Through Different Routes to Certification

Final Report

ies NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE
Institute of Education Sciences

An Evaluation of Teachers Trained Through Different Routes to Certification

Final Report

February 2009

Jill Constantine
Daniel Player
Tim Silva
Kristin Hallgren
Mary Grider
John Deke
Mathematica Policy Research, Inc.

Elizabeth Warner
Project Officer
Institute of Education Sciences

NCEE 2009-4043
U.S. DEPARTMENT OF EDUCATION

ies NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE
Institute of Education Sciences

U.S. Department of Education
Arne Duncan
Secretary

Institute of Education Sciences
Sue Betka
Acting Director

National Center for Education Evaluation and Regional Assistance
Phoebe Cottingham
Commissioner

February 2009

The report was prepared for the Institute of Education Sciences under Contract No. ED-01-CO-0039/0009. The project officer is Elizabeth Warner in the National Center for Education Evaluation and Regional Assistance.

IES evaluation reports present objective information on the conditions of implementation and impacts of the programs being evaluated. IES evaluation reports do not include conclusions or recommendations or views with regard to actions policymakers or practitioners should take in light of the findings in the reports.

To order copies of this report,

- Write to ED Pubs, Education Publications Center, U.S. Department of Education, P.O. Box 1398, Jessup, MD 20794-1398.
- Call in your request toll free to 1-877-4ED-Pubs. If 877 service is not yet available in your area, call 800-872-5327 (800-USA-LEARN). Those who use a telecommunications device for the deaf (TDD) or a teletypewriter (TTY) should call 800-437-0833.
- Fax your request to 301-470-1244.
- Order online at www.edpubs.org.

This report also is available on the IES website at http://ies.ed.gov/ncee.

Upon request, this report is available in alternate formats such as Braille, large print, audiotape, or computer diskette. For more information, please contact the Department's Alternate Format Center at 202-260-9895 or 202-205-8113.

ACKNOWLEDGMENTS

This study represents a collaborative effort of many schools, principals, program directors from teacher training programs, teachers, and researchers. We appreciate the willingness of principals, teachers, and administrative staff at study schools to provide access to classrooms and important data for the study, and the time principals and teachers spent in completing surveys and interviews. We also appreciate the time program directors spent providing detailed descriptions of their programs to research team members.

This report benefited from input from the technical work group: Dan Goldhaber, Tom Kane, Rob Hollister, Paul Holland, David Monk, Steve Rivkin, Jeff Smith, and Brian Stecher. Allen Schirm at Mathematica Policy Research, Inc. (MPR) provided critical technical review and comments. William Garrett led the production of the report.

The study would not have been possible without contributions from other individuals at MPR as well as our research partners, Decision Information Resources (DIR), Chesapeake Research Associates (CRA), and Vermont Institutes (VI). At MPR, Paul Decker was the project director for the first three years of the project, and Daniel Mayer was the deputy project director for the first year of the study. Amy Johnson led all aspects of the data collection effort, with assistance from Kathy Sonnenfeld. Martha Bleeker led the coordination of the observations of teachers and contributed to the analyses of teacher data. Nicole Saginor of VI led the training for classroom observations. The efforts to secure schools for the study and complete interviews with more than 80 directors of teacher training programs were particularly ambitious endeavors. We thank Nancy Dawson, Malené Dixon, Doug Hermond, Antwanette Hill, Jamie Liesmann, Ann McCoy, Scott Peecksen, Carla Prince, and Valerie Sheppard at DIR; Mike Puma and Dave Connell at CRA; and Nii Addy, Gail Baxter, Tim Bruursema, Jim Cashion, Scott Cody, Nancy Duda, Patricia DelGrosso, Benita Kim, Annette Luyegu, Jeffrey Max, Allison McKie, Melissa Miller, John Mullens, Debra Strong, Christina Tuttle, Cheri Vogel, Heather Zaveri, and Eric Zeidman at MPR for their professionalism, persistence, and good humor.

DISCLOSURE OF POTENTIAL CONFLICTS OF INTEREST[1]

The research team for this evaluation consists of a prime contractor, Mathematica Policy Research, Inc., of Princeton, New Jersey, and three subcontractors: Decision Information Resources (DIR), of Houston, Texas; Chesapeake Research Associates (CRA), of Annapolis, Maryland; and Vermont Institutes, of Montpelier, Vermont. None of these organizations or their key staff members have financial interests that could be affected by findings from the evaluation. No one on the Technical Working Group, convened by the research team to provide advice and guidance, has financial interests that could be affected by findings from the evaluation.

[1] Contractors carrying out research and evaluation projects for IES frequently need to obtain expert advice and technical assistance from individuals and entities whose other professional work may not be entirely independent of or separable from the tasks they are carrying out for the IES contractor. Contractors endeavor not to put such individuals or entities in positions in which they could bias the analysis and reporting of results, and their potential conflicts of interest are disclosed.

CONTENTS

Chapter **Page**

Chapter **Page**

IV *(continued)*

EXHIBITS

Exhibits **Page**

Exhibits **Page**

EXECUTIVE SUMMARY

Every year, thousands of new teachers pass through hundreds of different teacher preparation programs and are hired to teach in the nation's schools. Most new teachers come from traditional route to certification (TC) programs, in which they complete all their certification requirements before beginning to teach. In recent years, however, as many as a third of new hires have come from alternative route to certification (AC) programs, in which they begin teaching before completing all their certification requirements (Feistritzer and Chester 2002). AC programs have grown in number and size in recent years in response to a variety of factors, including teacher shortages and the No Child Left Behind (NCLB) Act, which requires that every core class be staffed with a teacher who has obtained full certification or, in the case of alternative routes to certification, is enrolled and making adequate progress toward certification through an approved program.

Despite the expansion of these new routes into teaching, there exists little research to provide guidance as to the effectiveness of different teacher training strategies. The increased variation in teacher preparation approaches created by the existence of various AC and TC programs offers an opportunity to examine the effect of different components of training on teacher performance. For example, some AC programs require less education coursework than TC programs. We can exploit this type of variation to examine whether the form of training is associated with differences in teacher performance.

The potential advantages and disadvantages of the various routes to certification have been debated, and the amount of coursework required by AC and TC programs is critical to issues of certification and teacher effectiveness. Some critics contend that the coursework required by TC (and some AC) programs is excessive and unnecessarily burdensome (Finn 2003; Hess 2001; U.S. Department of Education 2002), providing little benefit while discouraging talented people from entering the teaching profession (Ballou and Podgursky 1997). AC programs have been viewed as a way to eliminate these barriers. However, supporters of TC programs argue that easing requirements degrades quality because AC teachers are insufficiently prepared for the classroom and less effective than TC teachers (Darling-Hammond 1992). Even in cases where the coursework is similar, TC programs require that people complete their requirements prior to becoming a teacher of record, while AC programs allow them to begin teaching first. None of these claims, however, have been rigorously studied in the context of the programs that are most prevalent.

In light of these unresolved issues and the continuing need for highly qualified teachers, NCLB provides support "to ensure that teachers have the necessary subject matter knowledge and teaching skills in the academic subjects that the teachers teach." Specifically,

Title II of NCLB allows funds to be used for "carrying out programs that establish, expand, or improve alternative routes for state certification of teachers," as well as for "reforming teacher certification (including recertification) or licensing requirements." This study is intended to inform this effort by rigorously examining the effect of AC teachers on student achievement and classroom practices compared to the effect of TC teachers in their same school and grades. The study also provides suggestive evidence about what training and pretraining characteristics may be related to teacher performance.

Research on the effectiveness of AC teachers is not conclusive. A handful of studies have examined the effects on student achievement of specific AC programs, including Teach For America (TFA) and the New York City Teaching Fellows (NYCTF) program, and have reached mixed conclusions (Decker et al. 2004; Kane et al. 2006; Laczko-Kerr and Berliner 2002; Raymond et al. 2001). The more rigorous studies generally showed that students of AC teachers scored the same or higher than students of TC teachers, or that they scored slightly lower during their teacher's first year of teaching, but scored the same by the teacher's second year (Decker et al. 2004; Boyd et al. 2005; Kane et al. 2006). When effects have been found, they have typically been described by the authors as small. Some research—case studies or small-scale, nonexperimental observation and survey-based studies—has examined AC and TC teachers' classroom practices, and also had mixed findings (Lutz and Hutton 1989; Jelmberg 1996; Miller et al. 1998). Finally, because of their limited scope, many of these studies appear to have limited relevance to the broad range of AC programs operating across the country. The TFA and NYCTF programs, for example, recruit graduates from top colleges and are quite selective in admission, whereas the entry requirements of the majority of AC programs are less stringent (Walsh and Jacobs 2007; Mayer et al. 2003). Lacking conclusive evidence, principals may be uncertain of the implications of hiring an AC teacher, and policymakers may wonder about the implications of various characteristics of teacher certification programs.

RESEARCH QUESTIONS AND STUDY DESIGN

This study addresses two questions related to teacher preparation and certification routes:

1. What are the relative effects on student achievement of teachers who chose to be trained through different routes to certification? How do observed teacher practices vary by chosen route to certification?

2. What aspects of certification programs (such as the amount of coursework, the timing of coursework relative to being the lead teacher in the classroom, the core coursework content) are associated with teacher effectiveness?[2]

The answer to the first question is most relevant to principals faced with a choice between hiring an AC or a TC teacher. The answer to the second is of interest to

[2] Throughout the report, we use the terms "teacher effects" and "teacher effectiveness" to denote the effect of teachers on student achievement or classroom practices.

policymakers and designers and administrators of teacher training programs in their efforts to identify the training characteristics and certification requirements that are related most positively to student achievement.

A brief description of the study design is presented below, followed by a summary of the main study findings. More details on the selection of teacher preparation programs models, study sample, random assignment and analytical strategy, and data collection follow.

Study Design

Participants: Schools that had recently hired alternatively certified (AC) teachers were recruited to participate in the study. If the AC teacher was teaching the same grade level as a relative novice traditionally certified (TC) teacher, the school was eligible to participate in the evaluation. The evaluation included 2,600 students in 63 schools in 20 districts.

Research Design: In the study schools, every grade that contained at least one eligible AC and one eligible TC teacher was included. Students in these study grades were randomly assigned to be in the class of an AC or a TC teacher. The random assignment ensured that, within each teacher pair, the students in each classroom were similar on average. The pairing of an AC teacher to a TC teacher in each school and grade level constituted a separate mini-experiment. Students were tested at the beginning of the school year as a baseline measure and at the end of the year as an outcome. Classroom instruction was observed at one point during the year as an outcome.

Analysis: In each school grade, the outcomes of students who were randomly assigned to an AC classroom were compared to the outcomes of students who were assigned to a TC classroom, generating an impact estimate for each teacher pair, referred to as a mini-experiment. The overall impact was calculated by taking the average of the impacts from all mini-experiments. The mini-experiments were also divided into two approximately equal-sized subgroups based on the amount of coursework that was required (low or high) by the AC teacher's program, and the impacts were averaged separately for each group. Low-coursework AC teachers were defined as teachers whose program required 274 or fewer hours of coursework, while high-coursework AC teachers were defined as teachers whose program required 308 hours or more of coursework.

The main findings of the study are:

- ***Both the AC and the TC programs with teachers in the study were diverse in the total instruction they* required for their candidates.** The total hours required by AC programs ranged from 75 to 795, and by TC programs, from 240 to 1,380. Thus not all AC programs require fewer hours of coursework than all TC programs. The degree of overlap in coursework requirements between AC and TC programs in the study was dictated by variations in state policies on teacher certification programs. For example, in New Jersey all AC teachers were required to complete fewer hours of coursework than all TC

teachers, while in California, the range of coursework hours required was similar for AC and TC teachers.

- ***While teachers trained in TC programs receive all their instruction (and participate in student teaching) prior to becoming regular full-time teachers, AC teachers do not necessarily begin teaching without having received any formal instruction.*** Overall, low-coursework AC teachers in the study were required to take an average of 115 hours of instruction—64 percent of the total amount of instruction they would receive—before starting to teach, and high-coursework AC teachers in the study were required to take an average of 150 hours—about 35 percent of the total amount they would receive—before starting to teach. Nine AC teachers in the study, seven of them from New Jersey, were not required to complete any coursework before becoming regular full-time teachers.

- ***There were no statistically significant differences between the AC and TC teachers in this study in their average scores on college entrance exams, the selectivity of the college that awarded their bachelor's degree, or their level of educational attainment.*** Both low- and high-coursework AC teachers were more likely than their TC counterparts to identify themselves as black (40.5 percent versus 17.5 percent and 32.4 percent versus 7.5 percent) and less likely to identify themselves as white (50 percent versus 75.5 percent and 40.5 percent versus 70 percent). In addition, the low-coursework AC teachers were more likely than their TC counterparts to report having children (70.2 percent versus 28.3 percent).

- ***There was no statistically significant difference in performance between students of AC teachers and those of TC teachers.*** Average differences in reading and math achievement were not statistically significant. Furthermore, students of AC teachers scored higher than students of their TC counterparts in nearly as many cases as they scored lower (49 percent in reading and 44 percent in math). The effects of AC teachers varied across experiments, and nonexperimental correlational analysis of teachers' pretraining and training experiences explained 5 percent of the variation in math and 2 percent in reading. Therefore, the route to certification selected by a prospective teacher is unlikely to provide information, on average, about the expected quality of that teacher in terms of student achievement.

- ***There is no evidence from this study that greater levels of teacher training coursework were associated with the effectiveness of AC teachers in the classroom.*** The experimental results provided no evidence that students of low-coursework AC teachers scored statistically differently from students of their TC counterparts, nor did students of high-coursework AC teachers compared to those of their TC counterparts. Correlational analysis similarly failed to show that the amount of coursework was associated with student

achievement. Therefore, there is no evidence that AC programs with greater coursework requirements produce more effective teachers.

- ***There is no evidence that the content of coursework is correlated with teacher effectiveness.*** After controlling for other observable characteristics that may be correlated with a teacher's effectiveness, there was no statistically significant relationship between student test scores and the content of the teacher's training, including the number of required hours of math pedagogy, reading/language arts pedagogy, or fieldwork. Similarly, there was no evidence of a statistically positive relationship between majoring in education and student achievement.

Selection of Teacher Preparation Program Models

To provide information about effective methods of preparing and certifying teachers, the study design called for selecting a sample of teacher preparation models that were different from one another in structure and amount of coursework. Because the sampled programs were characteristic of the types of programs that train most of the nation's teachers, the study provides comparative information on teacher effectiveness for those able to hire from both routes. To shed light on whether the timing of training is related to the effect of teachers on student achievement and classroom practices, we focused on programs that place teachers in classrooms in one of two ways: (1) after the teachers have completed all their training (TC programs), and (2) before they have completed it (AC programs). In terms of coursework, we did not limit our focus within the pool of AC or TC programs, but for the analyses we distinguished the AC programs with relatively low coursework requirements from those with relatively high ones, which helped us assess whether increasing the volume of coursework is related to teacher effectiveness. Finally, all the AC programs in the study had to have less selective entrance requirements.[3] We focused on such AC programs for two reasons. First, most TC programs do not have highly selective entrance requirements (Hess 2001), nor do most AC programs (Walsh and Jacobs 2007; Mayer et al. 2003). Hence, less selective programs, whether AC or TC, are more policy relevant, since these are the programs that produce most teachers working today.

Second, AC programs with less selective entrance requirements are similar to the likely entrance requirements of the education programs attended by TC teachers in the study. To examine the relationship between preservice teacher training characteristics and teacher performance, it is important to disentangle the effects of the teacher training program on student achievement and classroom practices from the effects of pretraining teacher characteristics. Limiting the AC programs to the ones with entrance requirements similar to those of most TC programs helps to decrease at least some of the potential differences between teachers who attend AC or TC programs. For example, if the study included AC teachers entering through the TFA program or other highly selective teaching programs

[3] We defined "less selective" programs as those that did not require applicants to have a grade point average (GPA) in excess of 3.0.

who, on average, attended more selective undergraduate institutions and have higher SAT or ACT scores than teachers who attended less selective AC programs or TC programs, then it would be more difficult to determine whether relative differences in the classroom are due to the programs attended or to teachers' pretraining.

The Study Sample

The study sample was constructed, and the study was conducted, over two years. We began in late 2003 by identifying as many potentially eligible AC programs as possible. Among those states not known to have selective admissions criteria for their AC programs (12 total)[4] we compiled a list of 165 programs, from which we drew a random sample of 63, stratified to ensure diversity in terms of geography (state) and types of programs within states. For the 2004–2005 school year, we recruited schools that had hired teachers from a purposive subsample of the 63 sampled programs.[5] For the 2005–2006 school year, we sought more teachers from the same programs and also directly approached new districts in some of the same states that hired large numbers of AC teachers (for example, because they operated their own program). Schools could be included in the study only if they had at least one eligible AC and one eligible TC teacher in the same grade, in kindergarten through grade 5. To be eligible, teachers (1) had to be relative novices (three or fewer years of teaching experience prior to 2004–2005, five or fewer years prior to 2005–2006); (2) had to teach in regular classrooms (for example, not in special education classrooms); and (3) had to deliver both reading and math instruction to all their own students. The final study sample included 87 AC teachers and 87 TC teachers (some of whom participated in the study both years) from 63 schools in 20 districts and 7 states, as shown in Exhibit 1. Fourteen of the 20 districts were in urban areas, and 4 were on the fringe of one. Although we identified and sampled from a large number of less selective AC programs operating in 2003–2004, the programs and teachers that were included in the study sample were not necessarily representative of all AC programs operating at the time.

Random Assignment and Analytical Strategy

Within each school, students in the same grade were randomly assigned to either an AC teacher or a TC teacher. Each instance in which we conducted random assignment constituted a "mini-experiment"—achievement of students in a classroom taught by an AC teacher was compared to achievement of students in a classroom taught by a TC teacher. Because students in the classrooms were randomly assigned within the same school, the characteristics and motivations of students for each teacher pair[6] did not systematically

[4] We identified the 12 states based on available documentation, including various websites and Feistritzer and Chester (2002), and discussions with state education officials.

[5] We identified the subsample of programs through screening to ensure that the programs had at least one year of operational experience, would be in operation in the coming year, and had at least 12 graduates or enrollees teaching within a district.

[6] Each mini-experiment is a teacher pair, with a few exceptions: four mini-experiments involved three teachers, and two involved four teachers.

Exhibit 1. States, Districts, Schools, and Teachers in Study

State	Districts	Schools	AC Teachers	TC Teachers
California	5	15	20	18
Illinois, Wisconsin, Georgia, Louisiana	7	12	15	16
New Jersey	3	9	9	9
Texas	5	27	43	44
Total	**20**	**63**	**87**	**87**

differ, and the contextual situation was the same. This was done to minimize preexisting differences in students and schools that might influence teacher practices and student test scores. Thus the difference in student test scores can be attributed to the type of teacher and not student, classroom, or school characteristics. T-tests confirmed that there were no statistically significant differences in demographic characteristics, including gender, race/ethnicity, and eligibility for free or reduced-price lunch, or baseline achievement levels between students assigned to AC or TC teachers. In addition, the integrity of random assignment was well maintained: fewer than 3 percent of students originally assigned to one type of classroom switched over to the other type.

An important distinction of this design is that because certification routes are not randomly assigned to teacher trainees, the estimates of the effects on student achievement and classroom practices of teachers who were trained through different routes to certification pertain to those who *chose* to participate in these programs. Because of likely differences in the types of people who attend various certification programs, the results cannot be used to rigorously address how a graduate of one type of program would fare if he or she had attended another type. The study design and the collection of extensive data on teacher characteristics and experiences facilitate answering the second research question, concerning how student achievement and teacher practices are associated with teachers' training experiences toward initial certification. These findings are suggestive, however, because teachers were not randomly assigned to training programs or to their personal characteristics.

To estimate the effects of teachers who chose to be trained through different routes on student achievement and the classroom practices experienced by students, we compared teachers from AC programs with teachers in the same schools and grades who completed a TC program. We also estimated two subgroups—AC programs with low and high amounts of required coursework—to investigate separately the comparison of (1) AC teachers from low-coursework programs relative to their TC counterparts, and (2) AC teachers from high-coursework programs relative to their TC counterparts.[7] The comparison between AC and

[7] We determined which programs had low or high coursework requirements after interviewing their program directors, and the precise definitions are explained in Chapter III.

TC teachers overall provided an experimental estimate of the average difference in student achievement of teachers from the two routes, a comparison useful to principals and school administrators because it provides an indication of how students might perform when instructed by an AC teacher compared to a TC teacher. The subgroup estimates are of interest independent of the overall estimate, since there is variation in the amount of coursework required by state or district certification policy. The subgroup analyses allow us to determine, within an experimental framework, the effects on student achievement and classroom practices experienced by students of teachers who attended programs with a relatively large difference in required coursework as demonstrated by the comparison between teachers from low-coursework AC programs and their TC counterparts. We can also examine the effects on students of teachers who attended programs with relatively little difference in required coursework as demonstrated by the comparison between teachers from high-coursework AC programs and their TC counterparts.[8]

Data Collection and Measurement

Data for the study were collected from a variety of sources.

Student Achievement. We obtained information on students' reading and math achievement by administering the California Achievement Test, 5th Edition (CAT-5), published by CTB Macmillan/McGraw-Hill. See Appendix A for additional details.

Teacher Practices. We collected information on teachers' classroom practices in two ways. First, we directly observed and rated the quality of their instruction in literacy and math using the Vermont Classroom Observation Tool (VCOT), a proprietary instrument for classroom observations developed by the Vermont Institutes which covers three domains—lesson implementation, lesson content, and classroom culture. Second, we had principals rate the quality of the study teachers' reading/language arts instruction, math instruction, and classroom management relative to those of other teachers in the school. See Appendix A for additional details.

Teacher Characteristics. The main data source was a survey, administered in the spring, that collected information on teachers' professional backgrounds, the support they received during their first year as a full-time teacher, and their personal background characteristics. We also obtained their college entrance examination (SAT and ACT) scores.

Teachers' Certification Program Experiences. We interviewed program directors to collect detailed information on several major aspects of the training programs that study teachers attended, including the admission requirements, the amount of instruction required (overall and in five areas of particular interest designated by the study: classroom management, reading/language arts pedagogy, math pedagogy, student assessment, and child

[8] Low-coursework AC teachers were required to complete, on average, 179 hours of instruction, while their TC counterparts were required to complete an average of 671. High-coursework AC teachers were required to complete, on average, 432 hours of instruction, while their TC counterparts were required to complete 607.

development), the timing of instruction, the amount of required fieldwork, the length and features of student teaching assignments for TC teachers, and the provision of mentoring to AC teachers during their first year of teaching. The designation of AC teachers as either low-coursework or high-coursework, as well as measures of coursework in different subjects, reflects the requirements of the programs they attended and the amount of coursework required for certification, not the amount actually completed at the time of the study.

DESCRIPTIVE FINDINGS ON TEACHERS AND PROGRAMS

AC Teachers' Program Experiences

The AC teachers were required to take varying amounts of instruction in their programs, ranging from 75 to 795 hours. For analytical purposes, we divided AC teachers into two groups: the 47 who were required to complete 274 hours of instruction or less formed the low-coursework group, and the 40 who were required to complete 308 hours or more formed the high-coursework group. The low-coursework AC teachers' programs required an average of 179 hours of instruction (with a standard deviation [SD]of 54), while the high-coursework teachers' programs required, on average, 432 hours (SD of 112). Assuming that a typical college course involves about 45 hours of instruction (3 hours per week for 15 weeks), these means represent the equivalent of 4.0 and 9.6 courses, respectively.

Low- and high-coursework AC teachers also differed in the amount of coursework they were required to complete before, during, and after their first year of full-time classroom teaching, as shown in Exhibit 2.[9] For example, high-coursework AC teachers had to complete, on average, 150 hours of instruction during their first year of teaching, which translates to about 17 hours a month, compared with 63 hours, on average, among low-coursework AC teachers, which translates to about 7 hours a month.

TC Teachers' Program Experiences

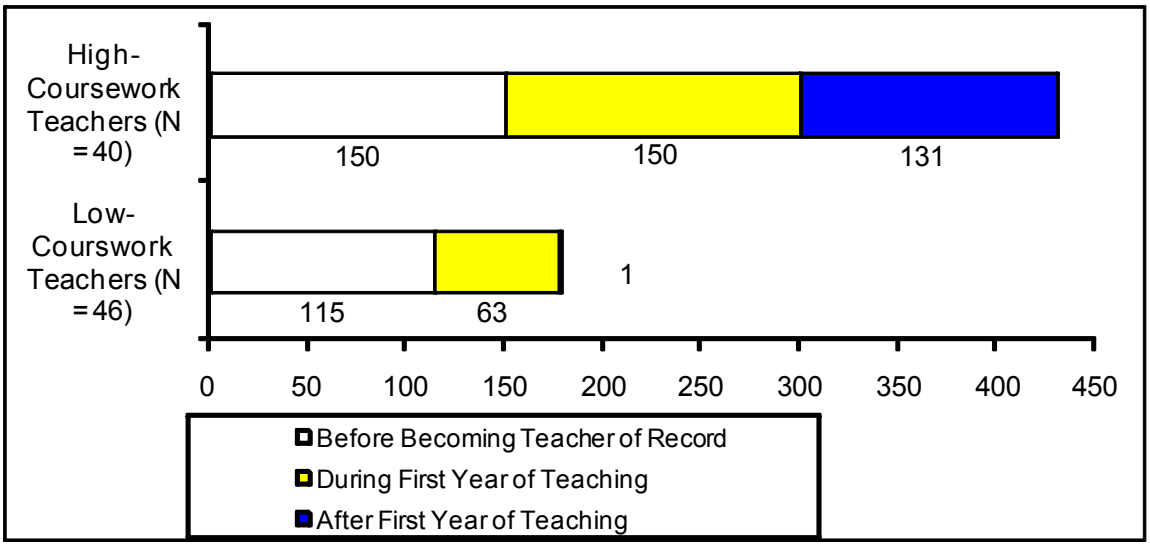
TC teachers, like their AC counterparts, received varying amounts of instruction, ranging from 240 to 1,380 hours. On average, they completed a total of 642 hours of instruction (SD of 225), equivalent to 14.3 typical college courses. This mean was more than double that of the AC teachers.

Comparisons of Instruction Required for AC and TC Teachers

We present data on four different groups of teachers: (1) teachers who chose low-coursework AC programs, (2) their TC counterparts, (3) teachers who chose high-coursework AC programs, and (4) their TC counterparts. In discussing the average amount

[9] One low-coursework AC teacher did not enroll in her program during the study year; therefore, we do not include required coursework hours for this teacher in Exhibit 2.

Exhibit 2. Average Hours of Instruction Relative to First Year of Teaching, AC Teachers



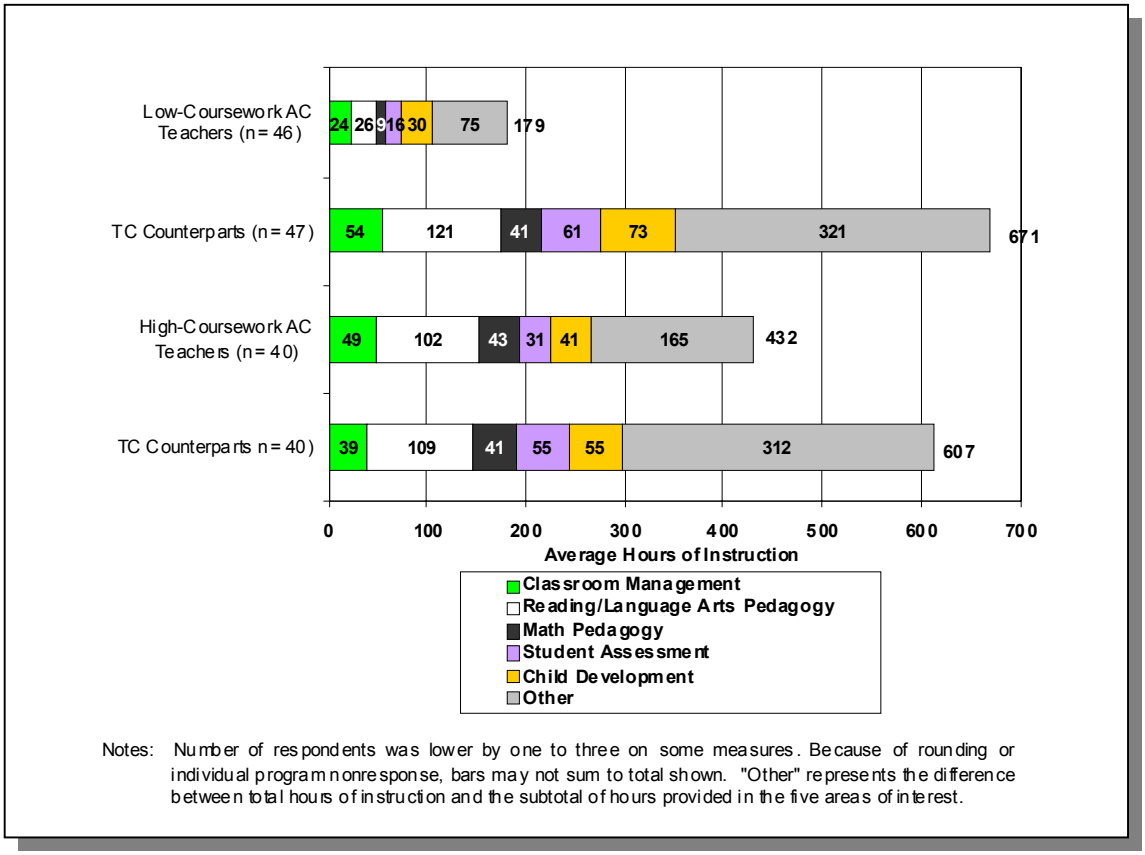
Source: Program director interviews.
Note: Because of rounding, bars do not sum to the averages reported earlier, 432 and 177.

of instruction that original study teachers were required to complete as part of their training programs, we examine differences between (1) the low- and high-coursework AC teachers, to explore the extent of differences in their programs' coursework requirements for certification; (2) the two groups of TC teacher counterparts to the low- and high-coursework AC teachers, to explore whether they provide a common benchmark for our experimental analyses[10]; and (3) each AC group and its counterpart TC group, to explore differences in coursework requirements that might be related to the results of the experimental and nonexperimental analyses presented below.

Coursework hours data collected for the study focused on five topics: reading/ language arts pedagogy, math pedagogy, classroom management, student assessment, and child development. We hypothesized that coursework hours in these specific topic areas would be most related to student achievement. However, because hours of instruction in topics other than these five accounted for 38 to 51 percent of the average total hours of required instruction for each group of teachers, we also discuss required hours of such instruction.

[10] If the two groups of TC teachers faced similar instructional requirements in their training programs, then both groups of AC teachers would face similar counterfactuals, and the key analyses (low-coursework AC teachers versus their TC counterparts, and high-coursework AC teachers versus their TC counterparts) would be comparable.

Exhibit 3. Average Hours of Instruction by Content Area, AC and TC Teachers



Notes: Number of respondents was lower by one to three on some measures. Because of rounding or individual program nonresponse, bars may not sum to total shown. "Other" represents the difference between total hours of instruction and the subtotal of hours provided in the five areas of interest.

Low- and High-Coursework AC Teachers. AC teachers from high-coursework programs were required to take more hours of instruction overall than AC teachers from low-coursework programs, as shown in Exhibit 3. As discussed above, dividing AC teachers into two similar-sized groups based on a gap in required coursework of AC programs yielded two groups with large average differences in required coursework. High-coursework AC teachers were required to complete 432 hours of instruction, compared with 179 for low-coursework AC teachers. This difference in total hours of instruction is due to differences in all five subject areas of interest as well as other instruction (defined below). High-coursework AC teachers were required to complete more hours of instruction in all five subjects, on average, than AC teachers from low-coursework programs: 3.9 times as much instruction in reading/language arts pedagogy, 4.8 times as much in math pedagogy, 2.0 times as much in classroom management, 1.9 times as much in student assessment, and 37 percent more in child development. Although not shown in Exhibit 3, all these differences were statistically significant at the 0.01 level, except for child development, which was statistically significant at the 0.05 level.

TC Teachers Matched to Low- and High-Coursework AC Teachers. TC teachers matched with low-coursework AC teachers were required to complete a similar amount of total instruction as TC teachers matched to high-coursework AC teachers, 671 hours versus 607, and the difference was not statistically significant. TC teachers matched with low-

coursework AC teachers were required to complete, in each of the five subject areas, on average, the same amount as or more instruction than TC teachers matched with high-coursework AC teachers, with statistically significant differences for classroom management and child development (at the 0.05 level; analysis not shown in Exhibit 3). Thus, in terms of required coursework, TC teachers matched to low- and high-coursework AC teachers served as a common benchmark in conducting the subgroup analysis.

Matched AC and TC Teachers Subgroups. AC teachers from low-coursework programs were required to complete, on average, about one-quarter of the total hours of instruction overall as their TC counterparts (179 hours versus 671 hours). In addition, they were required to complete less coursework in all subject areas of interest. For example, their programs required about one-fifth the instruction in reading/language arts pedagogy (26 versus 121 hours), less than one-fourth in math pedagogy (9 versus 41 hours), and less than half in classroom management (24 versus 54 hours). All the differences were statistically significant.

AC teachers from high-coursework programs were required to complete, on average, less instruction than their TC counterparts, 432 hours versus 607 hours, a difference that was statistically significant. They were required to complete less coursework in two topics of interest (student assessment, and child development), with the differences statistically significant. However, their programs required *more* instruction in classroom management (49 versus 39 hours), a difference that was statistically significant. There was no statistically significant difference in the amount of math pedagogy instruction (43 versus 41). Considering all five topics of interest together (that is, excluding "other" instruction), high-coursework AC teachers' programs required 91 percent as much instruction as their TC counterparts' programs (267 versus 295 hours), a difference that was statistically significant at the 0.05 level.

"Other" Instruction. For all teachers, some of the required coursework fell outside the five subjects of most interest in this study. Instruction in other topics accounted for, on average, 42 percent of total coursework for the low-coursework AC teachers, 48 percent for their TC counterparts, 38 percent for the high-coursework AC teachers, and 51 percent for their TC counterparts. "Other" instruction accounted for half the statistically significant 493-hour difference in total instruction between low-coursework AC teachers and their TC counterparts, and for 84 percent of the statistically significant 176-hour difference between high-coursework AC teachers and their TC counterparts.

AC and TC Teachers' Backgrounds

As context for interpreting the findings, Exhibit 4 presents information on the average background characteristics of the two AC teacher groups and their TC counterparts. Both low- and high-coursework AC teachers were more likely than their TC counterparts to identify themselves as black (40.5 percent versus 17.5 percent and 32.4 percent versus 7.5 percent) and less likely to identify themselves as white (50 percent versus 75.5 percent and 40.5 percent versus 70 percent). In addition, the low-coursework AC teachers were more likely than their TC counterparts to report having children (70.2 percent versus 28.3 percent). Low-coursework AC teachers had fewer years of teaching experience at the time

of their first year in the study, although the difference was less than one year. High-coursework AC teachers were more likely than their TC counterparts to be taking courses toward initial certification or an advanced degree during the study year (57 percent versus 30 percent). All these differences were statistically significant. Neither AC group had a statistically significant difference from its TC counterpart group in terms of college entrance exam scores or educational attainment.

Exhibit 4. Teacher Demographic and Educational Characteristics (Percentages, Except Where Noted)

	Low Coursework				High Coursework			
	AC	TC	Difference	*p*-Value	AC	TC	Difference	*p*-Value
White	48.8	73.8	–25.0	0.02	40.5	70.0	–29.5	0.01
Black	39.5	19.5	20.0	0.01	32.4	7.5	24.9	0.01
Female	95.7	97.9	–2.1	0.56	78.6	88.6	–10.1	0.21
Have children	70.2	27.7	42.6	0.00	38.1	29.5	8.5	0.41
Average age (years)	33.5	28.1	5.4	0.00	33.9	30.1	3.8	0.01
Average SAT or equivalent composite score[a] (points)	930	959	–29.0	0.43	1,010	1,013	–2.5	0.95
Highest degree: master's[b]	17.0	8.5	8.5	0.22	23.8	22.7	1.1	0.90
Currently taking courses[c]	31.9	21.3	10.6	0.25	57.1	29.5	27.6	0.01
Average study-eligible teaching experience (years)[d]	2.7	3.3	–0.6	0.04	3.3	3.0	0.2	0.45
Sample Size[e]	**46**	**46**			**42**	**44**		

Sources: Teacher survey for all but SAT scores, which were obtained from the College Board, and ACT scores, which were obtained from ACT.

[a]We converted ACT scores to SAT equivalents using the concordance procedure available from the College Board.

[b]All teachers had completed a bachelor's degree.

[c]Includes courses toward teaching certification or an advanced degree.

[d]Includes years teaching full-time as a certified or emergency certified teacher.

[e]Sample sizes were lower on some items due to nonresponse on the teacher survey; also, some teachers had not taken a college entrance exam, and others did not consent to release of their score. However, teachers who were in the study both years are counted twice here, whereas they were counted only once in earlier exhibits.

FINDINGS FROM EXPERIMENTAL ANALYSES

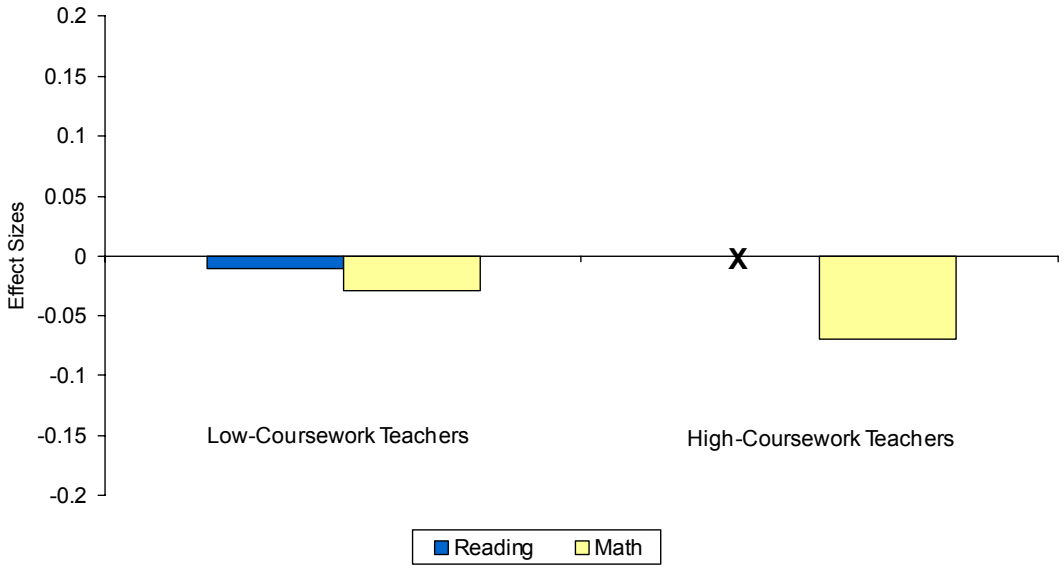
Students of AC teachers did not perform statistically differently from students of TC teachers. Although average differences in reading and math were generally negative, they were not statistically significant, as shown in Exhibit 5.

In addition to estimating the effects on student achievement of having a high- or low-coursework AC teacher, we examined effects within several subgroups to determine whether differences in teachers' effectiveness occurred within other dimensions even though differences did not exist overall. Specifically, we examined the relative effects of teachers in subgroups defined by state, current coursework status, grade level, and teaching experience.

All AC teachers in California were from high-coursework programs, and they accounted for half of all high-coursework AC teachers in the sample. Students of AC teachers in California scored lower on math than students of their TC counterparts, and the effect size (–0.13) was statistically significant. The effects of high-coursework AC teachers in other states was small (–0.01) and not statistically significant.

Students of AC teachers who were taking courses during the study year, toward either teacher certification or an advanced degree, had lower math scores than students of their TC counterparts (effect size = –0.09). The effect in reading was not statistically significant. Furthermore, neither the effect on reading nor the effect on math scores was significant for students of AC teachers who were not taking coursework during the study year.

Exhibit 5. Difference in Effect Sizes on Students' Reading and Math Scores of AC Teachers and Their TC Counterparts



Note: None of the effects was significantly different from zero at the .05 level.

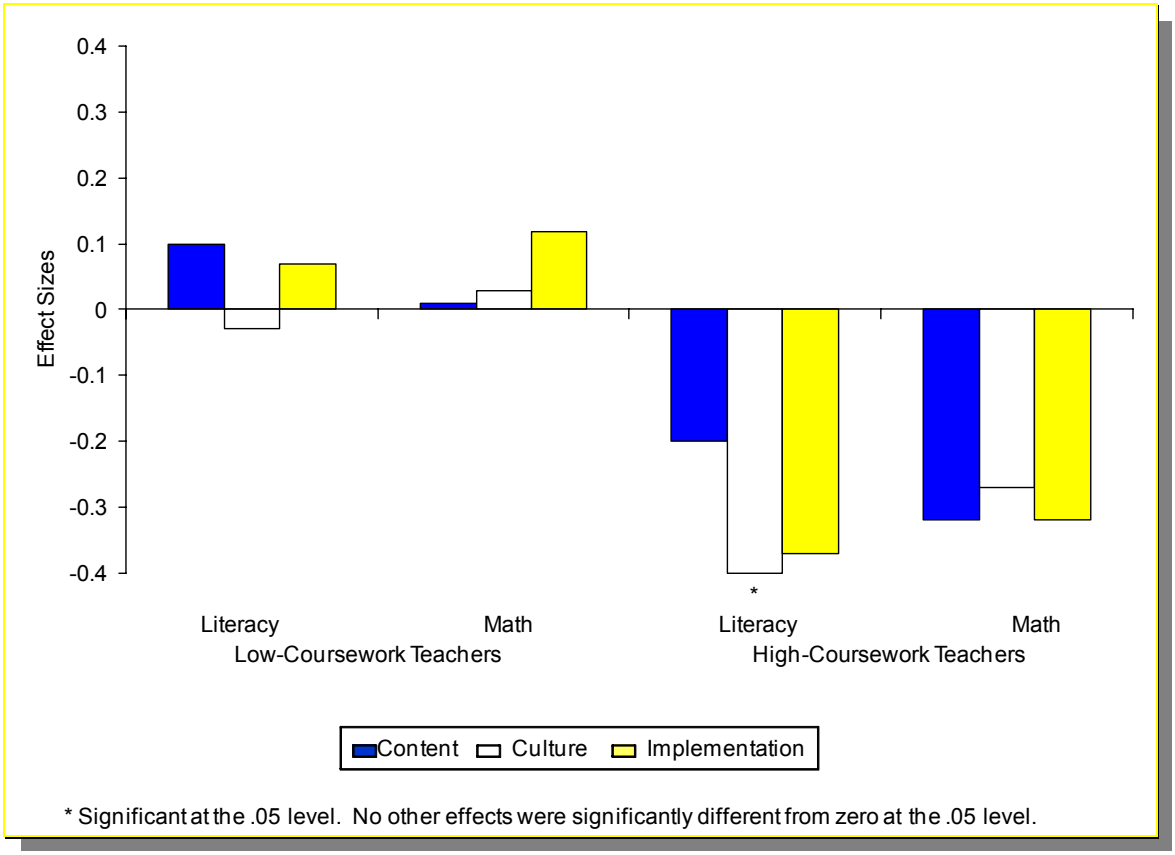
X The effect size was zero.

We found no evidence that AC teachers had a different effect on their students' math or reading achievement for different grade levels. There were no statistically significant differences between the lower elementary grades (K to 1) and the upper ones (2 to 5) for either the high- or the low-coursework AC teachers.

We found no evidence that students of AC teachers with less experience (1 to 2 years) had statistically significant different math or reading achievement, relative to their TC counterparts, than those with more experience (3 to 4 or 5 or more years). The one statistically significant difference pertained to students of low-coursework AC teachers in their third or fourth year of teaching, whose students scored lower in reading and math than students of their TC counterparts. Inferences based on these findings should be made with caution because the subgroup sizes were small and the experience levels of the TC comparison teachers varied.

With a single exception, ratings of classroom practices measuring the instruction received by students of AC and TC teachers did not differ. We found no statistically significant differences in VCOT scores between low-coursework AC teachers and their TC counterparts in the quality of their literacy and math instruction, as shown in Exhibit 6. High-coursework AC teachers also scored no differently from their TC counterparts on five of six VCOT measures, but they scored lower (by 0.40 SD) on the classroom culture dimension in teaching literacy, and the difference was statistically significant.

Exhibit 6. Difference in Effects Sizes on Classroom Practices of AC Teachers and Their TC Counterparts



* Significant at the .05 level. No other effects were significantly different from zero at the .05 level.

FINDINGS FROM NONEXPERIMENTAL ANALYSES

Although the average effect sizes (comparing achievement of students of AC teachers to achievement of students of their TC counterparts) were not statistically different from zero, effect sizes varied across individual pairs of AC and TC teachers. In reading, the effect size was less than zero in half the pairs and greater than zero in the other half. For math, the effect was less than zero in 56 percent of the pairs and greater than zero in 44 percent. Separating the effects of characteristics of teachers from the influences of their training, however, requires nonexperimental analysis, as does examining the relationship between teacher characteristics and classroom practices and student achievement.

To estimate the relationship between teacher characteristics and training experiences and student achievement, we used ordinary least squares (OLS) regression equations to estimate the correlation between a student's posttest score and student-level characteristics (including pretest score), whether his or her teacher was from an AC program, differences between the characteristics of AC and TC teacher pair within a school and grade, and other unobservable effects. This model allows us to estimate the relationship between differences in student achievement and differences in AC teachers and their TC counterparts' characteristics, such as required coursework, whether a teacher is currently taking courses, undergraduate major, and SAT scores.

All together, the differences in AC teachers' characteristics and training experiences explained about 5 percent of the variation in effects on math test scores and less than 1 percent of the variation in effects on reading test scores.

Differences in teachers' demographic characteristics and coursework required for initial certification were not related to the effects of teachers on student achievement. Of the several aspects of teachers' education and training we examined, two were statistically significantly related to the effects of teachers on student achievement, and both relationships were negative. First, AC teachers with master's degrees were less effective in improving student achievement in reading than their TC counterparts without a master's degree (effect size was –0.12). Second, students of AC teachers who were taking coursework toward certification or a degree scored lower in reading (effect size –0.13) than did students of their TC counterparts who were not taking coursework.

CONCLUSION

This study found no benefit, on average, to student achievement from placing an AC teacher in the classroom when the alternative was a TC teacher, but there was no evidence of harm, either. In addition, the experimental and nonexperimental findings together indicate that although individual teachers appear to have an effect on students' achievement, we could not identify what it is about a teacher that affects student achievement. Variation in student achievement was not strongly linked to the teachers' chosen preparation route or to other measured teacher characteristics.

CHAPTER I

INTRODUCTION

Over the past several decades, the U.S. labor market has experienced a growing shortage of teachers, largely because the potential supply has been reduced by improved opportunities for women (Corcoran et al. 2004; Stoddard 2003). At the same time, legislation aimed at reducing class size has increased the demand for teachers, particularly in schools that serve disadvantaged students (Hanushek et al. 2004).

Further, school districts are confronting these shortages in the face of increased pressure to hire only "highly qualified" teachers. One of the key provisions of the 2002 No Child Left Behind Act (NCLB), which was designed to reduce educational inequalities among students, is that every core class be staffed with a "highly qualified teacher," defined as one who "holds at least a bachelor's degree, has obtained full state certification, and has demonstrated knowledge in the core academic subjects he or she teaches" (U.S. Department of Education 2005). Although teacher certification is required by the law, the specific requirements for certification are decided by the individual states.

Increasingly, states are approving "alternative route to certification" (AC) programs that allow candidates to become a classroom teacher prior to completing all the requisite coursework and without having to complete a period of student teaching. In contrast, "traditional route to certification" (TC) programs require that candidates complete all coursework and a student teaching assignment before they begin teaching full-time.[11]

The potential advantages and disadvantages of the various routes to certification have been debated, and the amount of coursework required by AC and TC programs is critical to issues of certification and the effect of teachers on student achievement. Some critics contend that the coursework required by TC (and some AC) programs is excessive and unnecessarily burdensome (Finn 2003; Hess 2001), providing little benefit while discouraging

[11] Throughout the report we use "AC" to denote alternative routes to certification programs and "TC" to denote traditional routes to certification programs.

talented people from entering the teaching profession (Ballou and Podgursky 1997). AC programs have been viewed as a way to eliminate these barriers. However, supporters of TC programs argue that easing requirements degrades quality because teachers from AC programs are insufficiently prepared for the classroom (Darling-Hammond 1992).[12] Even in cases where the coursework is similar, TC programs require that people complete their requirements prior to becoming a teacher of record, while AC programs allow them to begin teaching first. None of these claims, however, have been rigorously studied in the context of the programs that are most prevalent.

In light of these unresolved issues and the continuing need for highly qualified teachers, NCLB provides support "to ensure that teachers have the necessary subject matter knowledge and teaching skills in the academic subjects that the teachers teach." Specifically, Title II of NCLB allows funds to be used for "carrying out programs that establish, expand, or improve alternative routes for state certification of teachers," as well as for "reforming teacher certification (including recertification) or licensing requirements." This study is intended to inform both types of efforts.

A. CONCEPTUAL FRAMEWORK FOR STUDY AND RESEARCH QUESTIONS

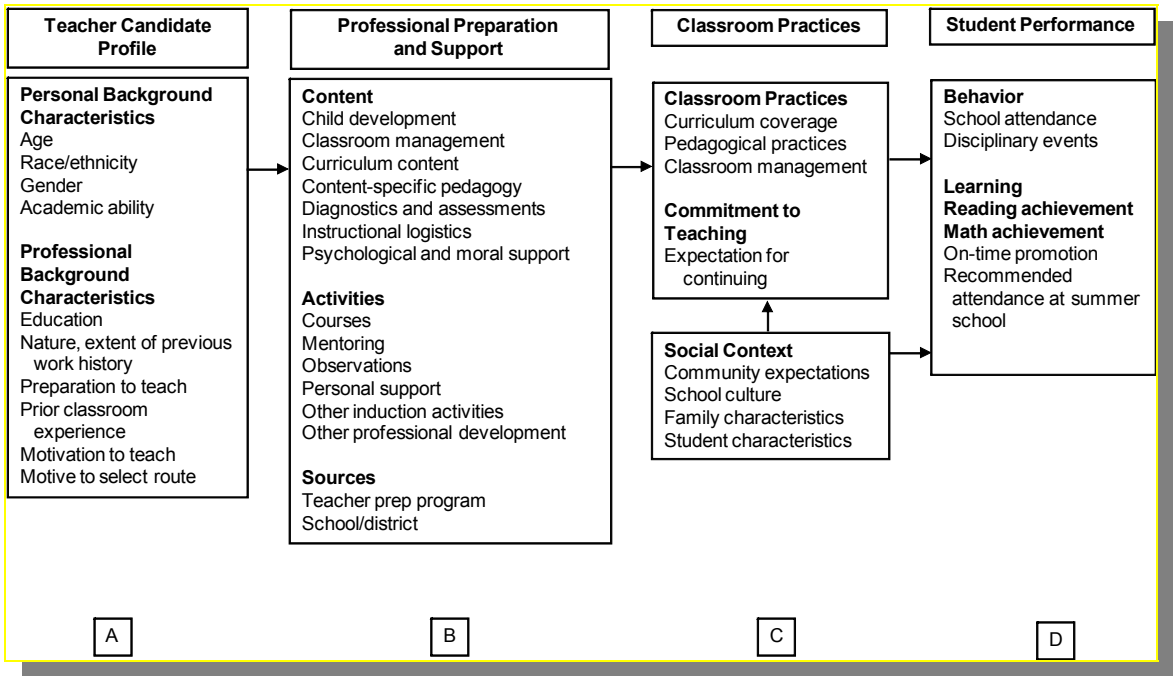
A conceptual framework for this study illustrates the potential contribution of preparation programs to teacher practices and student performance. This framework, depicted in Exhibit I.1, indicates core areas of exploration. It highlights the possible links between (1) teacher characteristics such as age, academic ability, education, and work experience (column A); (2) professional preparation and support during the early years of teaching (column B); (3) the intermediate effects these factors might have on classroom practices, which also are influenced by the social context (column C); and (4) the key longer-term effects that might be obtained on student performance, including school-related behaviors and achievement (column D).

The framework shows how both the pretraining characteristics of teachers and their preparation programs could be associated with classroom practices and student outcomes.[13] Describing the components of teacher preparation programs is a focus of this study, to provide context for interpreting the findings from analyses that address two major research questions. As explained below and in Chapter II, this study rigorously investigates the influence on student and teacher outcomes—by comparing results for teachers who attended two different types of AC programs with results for teachers who attended TC programs. The key dimensions on which the programs included in the study differ are the amount and timing of required coursework. The data also support nonexperimental analyses of the relationship between various characteristics and student and teacher outcomes.

[12] For ease of exposition throughout the report, we refer to teachers who obtained certification through AC programs as "AC teachers" and teachers who obtained certification through TC programs as "TC teachers."

[13] This study focuses on students' reading and math achievement outcomes. Other learning or behavior outcomes were not included, either because data were not consistently available or because the outcomes are rare in the grade levels included in this study (for example, disciplinary events).

Exhibit I.1. Conceptual Framework for Study of Teacher Preparation Models



This study addresses two research questions regarding teachers who have taken these different routes to certification:

1. What are the relative effects on student achievement of teachers who chose to be trained through different routes to certification? How do observed teacher practices vary by chosen route to certification?

2. What aspects of certification programs (for example, amount of coursework, timing of coursework relative to being the lead teacher in the classroom, core coursework content) are associated with teacher effectiveness?[14]

The answer to the first question is most relevant to principals and school administrators because it provides an indication of how students might perform when instructed by an AC teacher compared to a TC teacher. The answer to the second is of interest to policymakers

[14] Throughout the report, the terms "teacher effects" and "teacher effectiveness" are defined as the relative effect of teachers on student achievement as measured by a standardized achievement test or classroom practices as observed by trained, independent observers using the Vermont Classroom Observation Tool. Differences in teachers' pretraining characteristics, such as undergraduate major, achievement, and prior work experience, may vary by chosen route to certification and can also influence teacher performance. The report examines differences in pretraining characteristics along with characteristics of training programs.

and designers and administrators of teacher training programs in their efforts to identify the training characteristics and certification requirements that are related most positively to student achievement.

B. PREVIOUS RESEARCH

Every year, thousands of new teachers pass through hundreds of different preparation programs and are hired by our nation's schools, and as many as one-third of new hires are from AC programs (Feistritzer and Chester 2002). Along with the expansion of these new routes into teaching, several studies have examined teacher training programs and types of certification but have generally focused on specific programs or certifications. Thus, little empirical research exists to provide guidance as to the effectiveness of different teacher training strategies or to describe the characteristics of AC programs and the teachers they certify. In this section we present the previous research findings that motivate the questions addressed in this study. We also summarize findings from the few rigorous studies that have been conducted of AC teachers.

A key difference between AC and TC programs is the content and amount of required coursework, which could lead to differences in teacher effectiveness, as measured by classroom practices or student achievement.[15] A number of previous studies that have examined the relationship between the content of teacher training and student achievement have produced mixed results. There is nonexperimental evidence that students score higher in math if their teachers have taken more math classes in college (Monk 1994), but little evidence that there are benefits in other subjects, such as history, English, or science (Monk and King 1994; Goldhaber and Brewer 1997, 2000). In a correlational study of the effects of professional development and preservice teacher training, Harris and Sass (2007) find that preservice courses in pedagogical content knowledge are associated with positive returns in math test scores for elementary and middle school students, but they find no consistent evidence that increased coursework in educational theory, instruction, or class management is associated with improved student performance.

Another key difference between AC and TC programs is the timing of training: TC teachers complete their certification requirements before becoming classroom teachers, while AC teachers become teachers first. The little prior research on the relationship between the timing of training and student achievement suggests a negative effect when a teacher is taking coursework or completing certification requirements while teaching (Harris and Sass 2007; Goldhaber and Anthony 2006). One hypothesis is that the demands of coursework take time away from lesson preparation and other teaching-related duties.

[15] Differences in the content and amount of required coursework between AC and TC programs vary by state. In some states and districts there is no difference in the amount of required coursework, and AC programs might require the same or more coursework than TC programs in the same state or district. In other areas, AC programs require a fraction of the coursework required by TC programs (Walsh and Jacobs 2007). Differences in coursework requirements are discussed in more detail in Chapter III.

The average characteristics of teachers who enter through AC routes may also differ from those of the teachers who enter through TC routes. Previous nonexperimental research suggests that although teachers have a "powerful" effect on student achievement, very little of the effect can be explained by observable teacher characteristics, such as education, training, or experience (Rivkin et al. 2005, p. 417). While there is evidence that teacher experience is positively correlated with student outcomes, the relationship is most pronounced in the first several years of experience and tends to level off after that (Hanushek et al. 2005; Clotfelter et al. 2007). There is little evidence from nonexperimental research that advanced degrees are correlated with student achievement. In fact, research has often found a negative correlation between a master's degree and student achievement (Hanushek 1997; Clotfelter et al. 2007). Studies have shown that AC programs enroll a higher percentage of minorities, particularly African Americans, than TC programs (Zeichner and Schulte 2001; Peterson and Nadler 2009). While there is no evidence that the race of a teacher is related to student achievement in general, experimental and nonexperimental research shows either no effect or a positive and statistically significant effect on student achievement when African American students are matched to teachers of the same race (Ehrenberg and Brewer 1995; Ehrenberg et al. 1995; Dee 2004; Clotfelter et al. 2007).

A number of studies have examined the relationship between specific AC programs and student outcomes. The first set of studies examined Teach For America (TFA) (Decker et al. 2004; Laczko-Kerr and Berliner 2002; Raymond et al. 2001) and reached mixed conclusions. One study showed experimental evidence, and it concluded that students with TFA teachers scored the same in reading as students with comparable non-TFA teachers, and better in math (Decker et al. 2004). Two nonexperimental studies examined New York City AC programs, of which the New York City Teaching Fellows (NYCTF) program produces the most teachers. Using comprehensive data of students and their teachers, Kane et al. (2006) found that elementary students with NYCTF teachers scored approximately 0.01 standard deviation (SD) lower on their reading tests than students with TC teachers. Although this difference was statistically significant, it is equivalent to a difference of about two days of instruction. No statistically significant differences existed on math tests. The study also found that NYCTF teachers improve more after their first year of teaching than do TC teachers. Similarly, Boyd et al. (2005) found that students with NYCTF teachers in their first year of teaching scored 0.02 to 0.05 SD lower in reading and math than did students with TC teachers also in their first year. After the AC teachers gained two to three years of experience, their students scored at the same level as those of TC teachers. In a study of a third program, Miller et al. (1998) examined teachers from a single AC program in the southeast and found no statistically significant within-school differences between the achievement of the students in classrooms with AC as opposed to TC teachers.

In addition to the studies of AC teachers and student achievement, a number of case studies and small-scale, nonexperimental, observation- and survey-based studies have examined differences in classroom practices among AC and TC teachers. The results have been mixed. Miller et al. (1998) found no statistically significant differences in teaching behaviors between the two groups, while Lutz and Hutton (1989) and Jelmberg (1996) found that principals rated AC teachers lower than TC teachers on classroom practices.

The TFA and NYCTF programs are selective and recruit graduates of top colleges (Decker et al. 2004; Kane et al. 2006). However, AC programs serving the majority of teachers acquiring certification through alternative routes are characterized by less restrictive entry requirements, such as lower grade point averages (GPAs)[16] (Walsh and Jacobs 2007; Mayer et al. 2003). Likewise, the studies of teacher practices are limited largely to case studies or concentrated in a particular district or state. Therefore, the findings from prior research may have limited relevance for a broader class of AC programs and teacher training strategies.

C. CONTRIBUTIONS OF THE STUDY

This study makes several contributions to the research on the effectiveness of teachers trained through different preparation programs, including the following:

- ***A Study Sample That Includes Teachers from More AC Programs than Do Other Studies, Including In-Depth Information from a Representative Sample of Less Selective AC Programs from 12 Selected States.*** Although the study is limited to AC programs with less selective admissions criteria, these programs are more prevalent and have been less studied by researchers.

- ***Unbiased Estimates of Student Performance in AC and TC Teacher Classrooms in the Same Grade Levels at the Same Schools.*** Random assignment of students to teachers within schools ensures that estimates of the relative effects of teachers who choose AC and TC programs are unbiased and not confounded with preexisting student or school characteristics.

- ***An Analysis of How Variation in the Amount and Timing of Coursework Required in AC and TC Programs Relates to Student Achievement and Classroom Practices.*** The study takes advantage of the fundamental difference between AC and TC programs: the timing of required coursework. TC programs require that teachers complete all coursework before becoming the teacher of record for a class, and AC programs allow teachers to begin teaching while taking required coursework. In addition, the study uses state variation in coursework requirements for AC programs to compare student achievement and classroom practices experienced by students of teachers from relatively low-coursework AC programs to TC teachers who were required to complete, on average, more than three times more coursework hours. The study also compares teachers from relatively high-coursework AC programs to TC teachers who were required to complete, on average, one third more coursework hours.[17]

[16] We define "less selective" entry requirements for the purposes of this study more fully in Chapter III.

[17] The design of these analyses is described in more detail in Chapter II. The variation in coursework requirements is described in more detail in Chapter III.

D. LOOKING AHEAD

The rest of this report describes the study in detail and presents the findings from experimental and nonexperimental analyses. Chapter II describes the study design and analytical approach, the study sample, the data collection, and the characteristics of districts, schools, and students included in the study. Chapter III describes the teachers and programs. Chapter IV presents the results of the experimental and nonexperimental analyses conducted to address the research questions.

CHAPTER II

STUDY DESIGN AND DATA COLLECTION

The study design called for selecting a sample of teacher preparation programs that would best enable us to answer the research questions posed in Chapter I. In addition, it called for random assignment of students to teachers from AC programs and TC programs teaching in the same school and in the same grade, to ensure that estimates of the effect of teachers, as measured by student achievement and the classroom practices experienced by students, would be unbiased and not confounded with preexisting school or student characteristics. This chapter describes the rationale for specifying the types of teacher preparation programs included in the study, explains the study design and analytical approach, describes how the study sample was developed, presents data on the sample size and distribution, summarizes data collection, and describes participating districts, schools, and students.

A. TYPES OF TEACHER PREPARATION INCLUDED IN THIS STUDY

To address the study's research questions, we included teachers who selected routes to certification that differed from one another in structure and amount of coursework required. In terms of structure, we focused on two types of programs: (1) those that place teachers in classrooms only after they have completed teaching certification requirements (TC programs), and (2) those that place teachers in classrooms before they have completed their requirements (AC programs).[18] If we compared the performance of teachers from these two broad categories only, as measured by their students' performance on a standardized test, we would be testing the effectiveness of teachers who elected to complete all their certification requirements, including student teaching, before becoming a classroom teacher with the effectiveness of those who pursued a route that allowed them to begin teaching while completing their certification requirements.

[18] This functional definition of AC and TC programs was developed for the study and captures the fundamental distinction between the two routes to certification. In practice, however, individual programs within these categories may vary from the norm. For example, teachers from a TC program might receive waivers to become teachers of record before completing all their training, or teachers from an AC program might complete all their coursework before becoming teachers of record.

AC programs vary, however, on two key dimensions—the selectivity of their admission requirements and the amount of coursework they require (Mayer et al. 2003)—which allowed us to further refine our test of teachers who chose different routes to certification. Exhibit II.1 shows AC programs divided into four groups defined by these dimensions, as well as the two groups that were the focus of this study. In identifying AC programs whose teachers could be in the study, we ruled out those with more selective entrance criteria, which we defined as requiring applicants to have an undergraduate grade point average (GPA) of at least 3.0. We focused on AC programs with less selective entrance criteria for two reasons. First, most TC programs do not have highly selective entrance requirements (Hess 2001), nor do most AC programs (Walsh and Jacobs 2007; Mayer et al. 2003). Hence, less selective programs, whether AC or TC, are more policy relevant, since these are the programs that produce most teachers working today.

Second, AC programs with less selective entrance requirements are similar in terms of entrance requirements to the education programs attended by TC teachers in the study. In examining the relationship between preservice teacher training characteristics and teacher performance, it is important to disentangle the effects of the teacher training program on student achievement and classroom practices from the effects of pretraining teacher characteristics. Limiting the AC programs to those with entrance requirements similar to those of most TC programs helps to decrease at least some of the potential differences between teachers who attend AC or TC programs. Specifically, if the study design did not limit the AC programs included in the study sample to those with less selective admission requirements, there could be meaningful differences in the pretraining characteristics of the two groups of teachers studied. For example, if the study included AC teachers entering through the Teach For America program or other highly selective teaching programs who, on average, attended more selective undergraduate institutions and have higher SAT or ACT scores than teachers who attended less selective AC programs or TC programs, then it would be more difficult to determine whether relative differences in the classroom are due to the programs attended or to teachers' pretraining characteristics.[19]

Exhibit II.1. Alternative Certification Programs Included in the Study

		Entrance Requirements	
		Highly Selective	Less Selective
Coursework Required	Minimal		X
	Substantial		X

[19] While Chapter III examines differences in observable characteristics between AC and TC teachers in the study and Chapter IV controls for these differences in the correlational analysis, there is still the potential for differences in unmeasurable characteristics between teachers who chose to be certified through a less selective AC program instead of a TC program.

The other key dimension on which AC programs vary is the amount of coursework required to obtain certification. Policies dictating certification requirements for teachers who pursue alternative routes to certification vary by state and district. In some states, AC programs are mandated to include as many coursework hours or credits as TC programs, while in others, either coursework requirements are comparatively low or no specific amount of coursework is required. Since a key issue in the policy debate about teacher certification routes is whether teachers from AC programs are sufficiently prepared for the classroom, assessing whether the level of coursework required as part of teacher certification is related to student achievement is important for states and districts structuring AC programs.

B. STUDY DESIGN AND ANALYTICAL APPROACH

The study was designed to produce the most rigorous evidence possible in answering the first research question, concerning the relative effects on student achievement and teaching practices experienced by students of teachers who chose to be trained through different routes to certification. This objective required the use of random assignment. Specifically, students were to be randomly assigned to either an AC teacher or a TC teacher in the same grade, at the same school.[20] This design was selected to minimize preexisting differences in students and schools that might influence teacher practices and student test scores. Each instance of random assignment to either an AC or TC teacher within a school and grade constituted a "mini-experiment" (that is, a comparison between student outcomes and experiences for the AC teacher's classroom and the counterpart TC teacher's classroom).[21] Thus, the difference in student test scores can be attributed to the teacher and not student, classroom, or school characteristics. As we show later in this chapter, random assignment was successful in creating analytic groups that did not differ significantly on measurable characteristics; AC teachers and their TC counterparts were teaching similar students during the study. Since the random assignment was conducted within schools and grades, there were also no differences in school or grade characteristics.

An important distinction of this design is that because certification routes are not randomly assigned to teacher trainees, the estimates of the effects on student achievement and classroom practices of teachers who were trained through different routes to certification pertain to those who *chose* to participate in these programs. Because of likely differences in the types of people who attend various certification programs, the results cannot be used to rigorously address how a graduate of one type of program would fare if he or she had attended another type. The study design and the collection of extensive data on teacher characteristics and experiences facilitate answering the second research question, concerning how student achievement and teacher practices are associated with teachers' training experiences toward initial certification. These findings are suggestive, however,

[20] Details on the implementation of random assignment are in Appendix A.

[21] Throughout the report, we use the phrase "TC counterpart" to denote the TC teachers in the same schools and grades as the AC teachers in the study. Random assignment of students occurred for each AC teacher and the TC counterpart.

because teachers were not randomly assigned to training programs or to their personal characteristics.

To estimate the effects of teachers who chose to be trained through different routes on student achievement and the classroom practices experienced by students, we compared teachers from AC programs with teachers in the same schools and grades who completed a TC program. We also estimated two subgroups—AC programs with low and high amounts of required coursework—to investigate separately the comparison of (1) AC teachers from low-coursework programs relative to their TC counterparts, and (2) AC teachers from high-coursework programs relative to their TC counterparts.[22] The comparison between AC and TC teachers overall provided an experimental estimate of the average difference in student achievement of teachers from the two routes, a comparison useful to principals and school administrators because it provides an indication of how students might perform when instructed by an AC teacher compared to a TC teacher. The subgroup estimates are of interest independent of the overall estimate, since there is variation in the amount of coursework required by state or district certification policy. The subgroup analyses allow us to determine, within an experimental framework, the effects on student achievement and classroom practices experienced by students of teachers who attended programs with a relatively large difference in required coursework as demonstrated by the comparison between teachers from low-coursework AC programs and their TC counterparts. We can also examine the effects on students of teachers who attended programs with relatively little difference in required coursework as demonstrated by the comparison between teachers from high-coursework AC programs and their TC counterparts.[23]

There are two types of nonexperimental analyses, those using the subgroup comparisons described above and those using regression analyses examining whether differences in the characteristics of teachers' training programs or of teachers' pretraining were correlated with outcomes. Looking across the two subgroup comparisons, achievement for the students of teachers who chose certification through AC programs with low coursework requirements relative to their counterparts were compared to the achievement for students of teachers who chose certification through AC programs with high coursework requirements relative to their counterparts. A comparison of these subgroup findings enabled us to provide suggestive evidence about the degree to which the differential coursework requirements were associated with differences in student achievement and differences in teacher practices experienced by students. Second, we estimate multivariate regressions to estimate the relationship between teachers' characteristics and student achievement. The regression analyses included correlating the

[22] We determined which programs had low or high coursework requirements after interviewing their program directors, and the precise definitions are explained in Chapter III.

[23] Low-coursework AC teachers were required to complete, on average, 179 hours of instruction, while their TC counterparts were required to complete an average of 671 hours. High-coursework AC teachers were required to complete, on average, 432 hours of instruction, while their TC counterparts were required to complete 607.

amount and type of coursework required with student outcomes or teacher practices; and correlating teachers' pretraining characteristics, such as SAT or ACT scores, with outcomes.

C. THE STUDY SAMPLE

One objective in building a study sample was to find 90 total pairs of eligible AC and TC teachers, with about half the pairs including low-coursework AC teachers and half high-coursework AC teachers.[24] To be eligible, teachers had to be relatively new to the teaching profession, teach in regular classrooms (which excluded single-subject teachers, such as music, art, or special education teachers), and deliver reading and math instruction to their own students. A second objective was to include more AC programs in more states than had other studies, particularly programs with less selective admissions requirements, which, as explained above, produce most AC teachers and are therefore more policy relevant. We recruited schools and teachers for the study over two years, and our methods of finding suitable teacher pairs evolved during this period to enable us to meet our target sample size.

There is currently no comprehensive list of existing AC programs. Therefore, in late 2003, based on available documentation, including various websites and Feistritzer and Chester (2002), and discussions with state education officials, we identified 12 states that had active AC programs for elementary teachers and whose programs were not known to have selective admissions criteria. Some states, such as New York, were ruled out because of selective admissions requirements for programs statewide; many other states were ruled out for having no AC programs at all, or for having no AC programs training regular elementary teachers. From the 12 states, we obtained lists of all their AC programs training regular elementary teachers, 165 in all, ranging from one program each in Michigan and Wisconsin to 51 in California.

From the list of 165 programs we drew a stratified random sample of 63, which we refer to as the "representative sample" of AC programs in 12 selected states.[25] We then called these programs to assess their eligibility and suitability for the study. We verified that they used less selective admissions requirements, and then determined whether they had some operational experience (were not in their first year of operation; thus some of their recent candidates would have more than a year of teaching experience) and whether they would still be operating the next year (which ensured that officials would be available in the future to provide us with information about the programs). We inquired about the amount of

[24] This was the necessary sample size as indicated by our power analysis for detecting possible effects (Decker et al. 2005). The study was powered to detect effect sizes of approximately 0.15 for full sample analyses and approximately 0.20 for 50% subgroup analyses. For more details, see Appendix A.

[25] The states were divided into seven strata, based on number of programs and whether we understood their programs to have low or high coursework requirements: (1) California, (2) Colorado, (3) Texas, (4) New Jersey and Wisconsin, (5) Illinois and Michigan, (6) Arkansas and Louisiana, and (7) Georgia, New Mexico, and Tennessee. (We did not know for certain whether AC programs had low or high coursework requirements until we conducted detailed interviews with their directors, described later.) Within states, programs were grouped by type. For example, California's programs were of two types, those sponsored by institutions of higher education and those sponsored by school districts.

coursework they required of their candidates, and we gauged their potential as feasible sources for teachers we might include in the study. In particular, we sought programs that had at least a dozen recent enrollees or graduates teaching in schools within a district.

From the AC programs identified as meeting the eligibility criteria for the study, we requested lists of recent enrollees and the names of the elementary schools and districts in which they were teaching. We contacted schools during spring 2004 to explore the likelihood that a suitable teacher match would be in place for the following school year. To form a suitable match, the AC and TC teachers had to be relative novices (they could have up to three years of teaching experience), teach regular classrooms in the same grade level, and have self-contained instruction for reading and math. These efforts yielded a sample of 25 AC teachers and 24 TC teachers for the 2004–2005 school year, which required us to recruit again the following year to secure a sample large enough to detect possible effects.

For the 2005–2006 school year, we retained as many teachers as possible from the first year[26] and recruited additional teachers from the same programs who were teaching in the same schools or in other schools in the same districts. In addition, we directly approached new districts in some of the same states that hired large numbers of AC teachers (for example, because they operated their own program). Whenever we found teachers from AC programs not already represented in the study, we placed screening calls to ensure that the programs did not have selective admission standards. We also broadened the definition of "novice" teachers to include those with up to five years of teaching experience. This approach yielded a sample of 68 AC and 71 TC teachers in year two, which, when combined with the sample from year one, produced a total sample large enough to detect effect sizes on student achievement of 0.10 SD for the full sample and 0.20 SD for subgroups comprised of half the sample (Decker et al. 2005).

The final study sample is a purposive one, constructed to answer the study's research questions. The AC teachers (and TC teachers) included were not necessarily representative of everyone who attended the same training programs at the same time. Likewise, the programs that study teachers attended were not necessarily representative of all programs in the same categories (less selective AC programs with low coursework requirements, less selective AC programs with high coursework requirements, and TC programs). Finally, the study schools are not necessarily representative of the all schools that hire teachers from both AC and TC routes.

D. SIZE AND DISTRIBUTION OF STUDY SAMPLE

The final study sample included a total of 87 original AC and 87 original TC teachers from 63 schools in 20 districts and seven states, as shown in Exhibit II.2.[27] These counts of

[26] Fourteen teachers from year one of the study also participated in year two, with new classrooms of randomly assigned students. Six were AC teachers; eight were TC teachers.

[27] These totals count once the 14 teachers who were in the study both years.

Exhibit II.2. States, Districts, Schools, and Original Teachers in the Study

State	District	Schools	Teachers		
			AC	TC	Total
California	5	15	20	18	38
Illinois, Wisconsin, Louisiana, Georgia	7	12	15	16	31
New Jersey	3	9	9	9	18
Texas	5	27	43	44	87
Total	**20**	**63**	**87**	**87**	**174**

"original" teachers include those to whom students were randomly assigned and who were in a study classroom at the start of the school year following random assignment.[28] The final study sample is concentrated in Texas, which accounts for 50 percent of the original study teachers, and California, which accounts for 22 percent.

The final study sample included 90 mini-experiments, as shown in Exhibit II.3.[29] Most (84) involved random assignment of students to a pair of teachers, with one AC and one TC teacher. Four mini-experiments, however, involved three teachers (a "trio"), and two involved four teachers (a "quartet").[30] For analyzing outcomes, whenever a mini-experiment involved two AC or two TC teachers, we pooled data by teacher type.

Exhibit II.3 also reveals that a majority (56 percent) of the mini-experiments involved teachers at the two lowest elementary grade levels, kindergarten and grade 1. Since it is possible that classroom management and pedagogical approaches differ in the lower and upper elementary grades, we explore in Chapter IV the relative effects on student achievement and classroom practices of teachers in grades K to 1 separately from those in grades 2 to 5.

[28] Twelve original study teachers did not complete the school year in the classroom they were in at the time of random assignment. Information on departing teachers and their replacements is in Appendix A.

[29] There are more mini-experiments than original AC or original TC teachers because some teachers participated in both years of the study, as discussed earlier.

[30] Later in the report, we refer to all mini-experiments generically as teacher "pairs," regardless of the number of teachers included.

Exhibit II.3. Number and Structure of Mini-experiments, by Grade Level

	Structure of Mini-experiment				
Grade Level	Pair: 1 AC, 1 TC	Trio: 2 AC, 1 TC	Trio: 1 AC, 2 TC	Quartet: 2 AC, 2 TC	Total
K	18	1	1		20
1	27		2	1	30
2	14				14
3	9				9
4	10			1	11
5	6				6
Total	**84**	**1**	**3**	**2**	**90**

E. DATA COLLECTION AND MEASUREMENT

The conceptual framework presented in Chapter I guided the study's data collection plan. The framework indicated that to understand the relationship between teacher preparation programs and student and teacher outcomes, we needed information on students' backgrounds, teachers' backgrounds, school characteristics, and program characteristics. We also needed to measure those outcomes—students' achievement and teachers' classroom performance. Furthermore, to provide context for the findings, we collected information on teacher preparation programs more broadly, both those with teachers in the study and those from the 12-state sample of less selective AC programs. Here we provide an overview of the sources from which we collected data for the study. Additional details on data collection, along with information on response rates, are in Appendix A.

1. Data on Students in the Study

We obtained information on students' reading and math achievement by administering the California Achievement Test, 5th Edition (CAT-5), published by CTB Macmillan/McGraw-Hill. We conducted baseline testing in reading and math a few weeks after the start of each school year, and follow-up testing a few weeks before the end of the school year. We also collected school records with information on each student's gender, race/ethnicity, and eligibility to receive a free or reduced-price lunch (FRPL).

2. Data on Teachers in the Study

Classroom Practices

We collected information on teachers' performance in the classroom in two ways. First, we observed teachers teaching two regular literacy (reading and writing) lessons and two regular math lessons and scored their performance using the Vermont Classroom Observation Tool (VCOT), a proprietary instrument for classroom observations developed by the Vermont Institutes. The VCOT covers three domains, each based on four to seven

separate indicators of what the instrument's developers believe to be good teaching practices.[31] The *lesson implementation* domain measured the use of best practices, pacing, teacher confidence, and student engagement. The *lesson content* domain measured the teacher's understanding of the concepts and content of the lesson, applicability of content and class assignments to the real world, and connections to other subjects or lessons. And the c*lassroom culture* domain measured the clarity and consistency of classroom routines, respectfulness and appropriateness of behavior, and teacher sensitivity to student diversity. For each indicator, trained observers scored teachers as having shown (1) no evidence, (2) limited evidence, (3) moderate evidence, (4) consistent evidence, or (5) extensive evidence. We calculated average scores at the domain level.

Second, using a form we developed for the study, principals rated how well each study teacher performed relative to all other teachers in the school. The form contained four indicators of the quality of reading/language arts instruction, four of the quality of math instruction, four of the quality of classroom management skills, and one of how well the teacher utilized parents and school resources. Ratings were on a 5-point scale, where 1 = substantially below average, 3 = average, and 5 = substantially above average. In our analysis of these data, we averaged the ratings for each teacher on the four reading/language arts items, the four math items, and the remaining five items.

Background Characteristics

The main source of information on the characteristics of study teachers was a survey that collected information on educational and professional backgrounds, types of support received during the first year as a full-time teacher, and personal background characteristics. We also collected information on the selectivity (admissions competitiveness) of the teachers' undergraduate institution from *Barron's Profiles of American Colleges 2003*. Finally, for teachers who had taken either the SAT or ACT college entrance examinations and who gave us written consent, we obtained the scores from the respective sponsors, the College Board and ACT. We converted ACT scores to SAT equivalents using concordance tables available from the College Board.

[31] The VCOT math observation tool was based on the quality of standards-based, investigative science and mathematics instruction, created by Science and Math Program Improvement (SAMPI), a research group at Western Michigan University, and based on research conducted by Horizon Research, Inc. It was further refined by staff at Vermont Institutes based on Charlotte Danielson's Framework for Teaching (1996). The VCOT literacy observation tool was based on the math tool but adapted to incorporate the standards and practices included in the National Council of Teachers of English Standards and the National Reading Panel (NICHD 2000). For more details, see Appendix A.

Certification Program Experiences

Through interviews with program directors, we collected detailed information on five aspects of the training programs that study teachers attended: (1) the admission criteria; (2) the amount, timing, and content of the instruction required; (3) the amount of fieldwork required; (4) the features of student teaching assignments for TC teachers; and (5) the nature of mentoring provided to AC teachers during their first year as a teacher of record.

- ***Admission criteria*** refers to requirements for admission to a teacher training program. These included minimum overall GPA or GPA in specific courses, letters of recommendation, and an interview with program staff.

- ***Instruction*** refers to time that candidates are required to spend in class with an instructor in lectures and seminars—"seat" or "contact" time—as well as on structured, self-paced assignments, such as computer-based tutorials.[32] We determined the total clock hours of instruction required for each program, and obtained estimates of the hours required in five areas that we hypothesized could influence the study's main outcomes: (1) classroom management, (2) reading/language arts pedagogy, (3) math pedagogy, (4) student assessment, and (5) child development. Instruction in each of these topics could have been the focus, or just a part, of one or more full courses. Any hours of instruction not counted toward one of the five areas of interest was counted as "other."

- ***Fieldwork*** refers to time that candidates are required to spend in elementary or middle school classrooms observing teachers and students, working with students, or leading lessons. It does not include student teaching.

- ***Student teaching*** refers to time that TC teachers spend in a local school under the tutelage of a regular classroom teacher, as the culmination of their training program. We collected information on the hours devoted to student teaching each day, the length of the experience in weeks, the number of full-length school days that student teachers were expected to spend fully in charge of their classrooms, the number and length of student teacher observations conducted by program-based field supervisors, and the number and length of seminars or other meetings associated with student teaching.

- ***Mentoring*** refers to personal one-on-one support provided by AC programs to teachers during their first year of teaching. We collected information on

[32] This does not include time spent on unstructured independent study or on preparing for tests or completing course assignments.

whether a mentor was provided to the teacher, and the type and frequency of mentoring activities.[33]

3. Data on a Representative Sample of Less Selective AC Programs in 12 States

To provide a comparative context for considering the characteristics of the AC programs attended by study teachers, we collected information on less selective elementary AC programs in the representative sample from 12 selected states described earlier. Screening calls revealed that 9 programs were no longer in operation, which reduced the sample from 63 to 54. Interviews concerning the remaining programs addressed admission requirements, amount and timing of instruction, and mentoring, while focusing on the programs as they existed between 2001 and 2004, the period when most AC study teachers were enrolled in their programs. These interviews, however, were less detailed than the ones we conducted concerning the programs with teachers in the study.

4. Data on Schools and Districts in the Study

We used the 2004–2005 and 2005–2006 Common Core of Data (CCD), a public database available from the U.S. Department of Education, to collect information on the schools and districts in our study sample. The CCD includes information on school enrollment, the percentage of students who are nonwhite, and the percentage of a school's students eligible for FRPL. It also contains data on school staff, but not on the type of training program teachers attended, so we relied on principals for that information. During interviews at study schools, we asked principals to report the total number of regular, self-contained classroom teachers in grades K to 5 at their school, and the number of these teachers who were currently enrolled in or who had been certified through an AC program.[34]

F. Characteristics of Districts, Schools, and Students in the Study

Since this study uses a purposive sample, we describe the districts, schools, and students in the sample to provide a context for understanding the settings and students for which the study findings are most relevant.

[33] We did not gather data on mentors from TC program directors, because TC teachers are generally not provided mentors from their *programs* during their first year of teaching. Rather, both AC and TC teachers may be provided with a *school-* or *district-based* mentor during their first year of teaching; we asked about this experience in the teacher survey.

[34] We focused on regular, self-contained classroom teachers, because only such teachers were eligible for the study. About 25 percent of principals indicated uncertainty about the AC/TC status of teachers at their schools, but gave us their best estimates nonetheless.

1. Characteristics of Districts in the Study

Of the 20 districts in the study, 14 were in urban areas and 4 were in the fringe of an urban area (Exhibit II.4). One district was in a town, according to the CCD, and one was in

Exhibit II.4. Characteristics of Districts in the Study

	District	Locale	Number of Elementary Schools in District	Total Elementary Enrollment	Percentage of Students Eligible for Free or Reduced-Price Lunch[a]	Percentage of Students Nonwhite[a]
California	District A	Urban	10	5,439	26	91
	District B	Urban	24	16,693	95	100
	District C	Urban	473	365,256	77	89
	District D	Urban	63	24,568	62	91
	District E	Urban	132	64,664	64	75
Georgia	District F	Urban	29	13,822	78	80
Illinois	District G	Urban fringe	10	2,937	85	92
	District H	Urban fringe	7	2,369	79	99
Louisiana	District I	Urban	12	5,166	84	91
	District J	Town	11	3,278	78	65
	District K	Rural area	29	11,359	76	54
New Jersey	District L	Urban	22	10,372	85	99
	District M	Urban fringe	21	12,728	78	89
	District N	Urban	52	27,170	47	95
Texas	District O[b]	Urban	32	25,061	n.a.	99
	District P	Urban	146	92,981	87	95
	District Q	Urban fringe	37	27,371	34	72
	District R	Urban	198	115,969	81	93
	District S	Urban	28	17,993	59	65
Wisconsin	District T	Urban	117	48,234	81	82

Source: CCD 2004–2005, except for districts E, L, M, and N, for which we relied on CCD 2005–2006.

[a]Average percentage across the district's elementary schools; variable not available in CCD at district level.

[b]The CCD reported that no students were eligible for FRPL in any of the elementary schools in this district. We believe this to be incorrect, as FRPL rates are non-zero for higher-level schools, and school records showed that 95 percent of District O students in the study were eligible for FRPL.

n.a. = not available.

a rural area. This distribution reflects our recruitment strategy of focusing on areas with concentrations of AC teachers. The smallest district had seven elementary schools serving just fewer than 2,400 students, while the largest had 473 elementary schools serving just over 365,000 students. Five districts served fewer than 10,000 elementary students, 11 had between 10,000 and 50,000, two had between 50,000 and 100,000, and two had more than 100,000. In 13 districts, elementary schools had an average of more than 75 percent of their students eligible for FRPL, and in 16 districts, the average proportion of nonwhite students was also greater than 75 percent.

2. Teaching Staff of Study Schools

The 63 schools in the study had an average of 35 regular classroom teachers in grades K to 5, with a range of 9 to 75. The number of teachers in these schools who were from AC programs averaged 8, with a range of 1 to 33. At 38 percent of the schools, at least a quarter of the regular classroom teachers in grades K to 5 were from AC programs, and at four schools, AC teachers accounted for more than half of all the teachers. The representation of AC teachers in study schools is an artifact of the study design and our approach to building the sample; as explained earlier, schools had to have at least one novice AC teacher in grades K to 5 to be eligible for the study, and we targeted districts that expected to hire teachers each year from one or more AC programs.

3. Other Characteristics of Study Schools

Using the CCD, we examined whether study schools were similar to non-study elementary schools in the same districts on two demographic measures: eligibility for FRPL and race/ethnicity.[35] The schools in the study were more economically disadvantaged than non-study schools in the same districts. In 14 of the 20 districts, the average percentage of students eligible for FRPL in the study schools was greater than the average percentage in non-study elementary schools, and in 8 cases the difference was 10 percentage points or more (Exhibit II.5). In 5 districts, the average percentage of students eligible for FRPL in the study schools was lower than the average percentage in non-study schools, and the difference was more than 10 percentage points in 1 case. The study schools also had higher minority enrollments than non-study schools in the same districts. In 16 districts, the average percentage of nonwhite students in the study schools was greater than the average percentage in non-study schools, and in 4 cases the difference was 10 percentage points or more (Exhibit II.5). In 2 of 20 districts, the average percentage of nonwhite students in the study schools was lower than the average percentage in non-study schools, and in neither case was the difference more than 10 percentage points.

4. Students' Baseline Characteristics

Average baseline test scores for the four groups of students were lower than national averages (below 50 normal curve equivalents, or NCEs), as shown in Exhibit II.6, but within 12 NCEs of the national average. Students of high-coursework AC teachers scored, on average, 38.44 NCEs on reading and 41.34 NCEs on math in the fall of the year they were in the study, compared with 37.99 and 42.63 for students of their counterpart TC teachers. Students of low-coursework AC teachers scored, on average, 39.88 on reading and 43.46 on math, compared with 39.93 and 43.05 for students of their TC counterparts. None of the differences in baseline test scores were statistically significant.

[35] The data reported in Table II.5 are based on all schools in each district. Thus all differences reported are true differences, and we do not report on the statistical significance of the differences.

Exhibit II.5. Average Characteristics of Study Schools and Non-study Schools, by District

		Percentage of Students Eligible for Free or Reduced-Price Lunch		Percentage of Students Nonwhite	
	District	Study Schools	Non-study Elementary Schools	Study Schools	Non-study Elementary Schools
California	District A	27	28	94	90
	District B	94	95	100	99
	District C	92	77	98	88
	District D	68	62	96	90
	District E	95	63	98	72
Georgia	District F	95	76	96	77
Illinois	District G	91	84	99	91
	District H	87	77	100	99
Louisiana	District I	97	83	99	90
	District J	86	76	88	60
	District K	95	75	92	52
New Jersey	District L	87	85	99	99
	District M	80	77	93	88
	District N	35	48	89	95
Texas	District O[a]	n.a.	n.a.	99	99
	District P	95	87	99	95
	District Q	33	35	73	72
	District R	87	81	99	93
	District S	54	61	60	66
Wisconsin	District T	98	80	88	82

Source: CCD 2004–2005, except for districts E, L, M, and N, for which we relied on CCD 2005–2006.

[a]The CCD reported that no students were eligible for FRPL in any of the elementary schools in this district. We determined that this was an error for the study school, however, as 95 percent of District O students in the study were eligible for FRPL, according to school records we collected. We believe the data on the non-study schools are also incorrect, as FRPL rates are non-zero for higher-level schools.

n.a. = not applicable.

Students in the study tended to come from poor families and to be racial/ethnic minorities. Average poverty rates, indicated by eligibility for FRPL, ranged from 65 to 84 percent among the four groups of students, and the average percentage of students who were racial/ethnic minorities ranged from 85 to 95 percent. However, as with baseline test scores, the differences between analytic groups were not statistically significant, which indicates that random assignment produced statistically equivalent groups of students.

Exhibit II.6. Average Baseline Characteristics of Students in AC and TC Classrooms

	AC Classrooms	TC Classrooms	Difference	*p*-Value
Pairs Involving High-Coursework AC Teachers				
Reading pretest score[a]	37.63	36.86	0.78	0.45
Math pretest score[a]	40.28	41.70	–1.29	0.26
Eligible for free/reduced-price lunch	84%	86%	–2%	0.26
Male	47%	46%	1%	0.82
Nonwhite	95%	95%	0%	0.93
Sample Size[b]				
Teachers	42	45		
Students	598	681		
Pairs Involving Low-Coursework AC Teachers				
Reading pretest score[a]	39.91	39.28	0.63	0.54
Math pretest score[a]	43.47	42.51	0.96	0.38
Eligible for free/reduced-price lunch	67%	69%	–2%	0.16
Male	46%	43%	3%	0.32
Nonwhite	88%	87%	1%	0.48
Sample Size				
Teachers	51	50		
Students	678	653		

Source: CAT-5, administered by Mathematica Policy Research, Inc., and administrative records. The math test refers to the Mathematics Concepts subsection.

Note: Sample sizes indicate the final sample with non-missing posttest scores. Missing values are imputed on variables other posttest scores.

[a]Test scores are expressed in terms of NCEs; the average score nationally is 50, and the SD is 21.06.

[b]Sample sizes shown are for reading; for math, sample sizes were lower because one teacher pair that did not teach math was omitted. Math sample sizes were 41 AC classrooms with 582 students and 44 TC classrooms with 666 students.

CHAPTER III

TEACHERS AND PROGRAMS IN THE STUDY

This chapter focuses on the study teachers and their training programs to provide a context for considering the results of analyses presented in Chapter IV. First, we describe the AC teachers and the programs they attended. Second, we compare these programs with a representative sample of less selective elementary AC programs in 12 selected states. Third, we describe the TC "counterparts," that is, the teachers teaching in the same school and grades as the AC teachers, and the programs they attended. Finally, we compare AC and TC teachers on the amount of instruction and fieldwork their programs required and on their background characteristics and professional experience.

A. CHARACTERISTICS OF AC TEACHERS AND THE PROGRAMS THEY ATTENDED

This section provides background on the training experiences of AC teachers in the study and the programs they attended. We first identify the sponsors of the AC programs with teachers in the study and highlight the different types of sponsoring organizations. Second, we present information on the total amount of instruction that programs required for their candidates and explain how we distinguished AC programs with low-coursework requirements from AC programs with high-coursework requirements.[36] Third, we describe the amount of coursework for certification that teachers had to complete before, during, and after their first year of teaching for certification. Fourth, we describe how AC programs mentor and support their participants during their first year as full-time teachers.

1. Sponsoring Organizations

The 87 AC teachers in the study attended programs sponsored by 28 organizations across seven states (Exhibit III.1). Sixteen of the sponsoring institutions were colleges or universities, half of which also operated TC programs whose graduates were in the

[36] Information on the amount of instruction required in certain topics and on the amount of fieldwork required is in Section D, where the data for AC teachers can be directly compared to the data for TC teachers.

Exhibit III.1. Sponsors of Programs Attended by Original AC Teachers in Study, by State of Teaching Assignment

State	Sponsor
California	California State University, Dominguez Hills[a] California State University, East Bay (Hayward) California State University, Los Angeles California State University, San Marcos Compton Unified School District Los Angeles Unified School District[a] San Diego Unified School District[a] San Jose State University University of California Berkeley Extension
Georgia	Wesleyan College and Bibb County Public Schools[a]
Illinois	Governors State University[a]
Louisiana	Louisiana College[a] Northwestern State University University of Louisiana, Monroe[a]
New Jersey	Elizabeth Regional Training Center Kean University[a] Montclair State University New Jersey City University St. Peter's College[a]
Texas	ACT Houston[a] Alternative Certification for Teachers in the Rio Grande Valley Dallas Independent School District[a] Houston Independent School District[a] Mountain View Community College Region 4 Education Service Center Region 12 Education Service Center Tarleton State University[a]
Wisconsin	Milwaukee Public Schools
Total	**28**

Source: Teacher self-reports on a study eligibility form.

Note: All program sponsors were located in the state under which they are listed. None of the AC teachers in the study were teaching in a state different from the one where they had received, or were receiving, their training.

[a]These sponsoring institutions had programs that were part of the representative sample described in Chapter II. One sponsor, University of Louisiana, Monroe, had two AC programs in the representative sample.

study.[37] Six program sponsors were school districts that were using the program as part of a strategy to help meet their own schools' needs for new teachers, either for all kinds of elementary teachers or for certain kinds, such as bilingual specialists. The remaining six AC program sponsors included private nonprofit organizations (such as ACT Houston), providers of education services to multiple districts or regions within a state (such as Texas's regions 4 and 12 Education Service Centers), and consortia of different types of organizations (for example, Bibb County Public Schools and Wesleyan College jointly operated a program in Georgia).

As will be seen below, the training experiences of the 87 AC teachers in the study were more diverse than may be suggested by the fact that they attended programs sponsored by 28 organizations. This is because some sponsors operated multiple training programs with different features, such as one for bilingual teachers and one for English-only teachers, or one for early childhood teachers and one for other elementary teachers. Furthermore, even teachers who attended the same program could have different training experiences, because the program requirements may have changed over time.

2. Total Hours of Instruction: Distinguishing Low- and High-Coursework AC Teachers

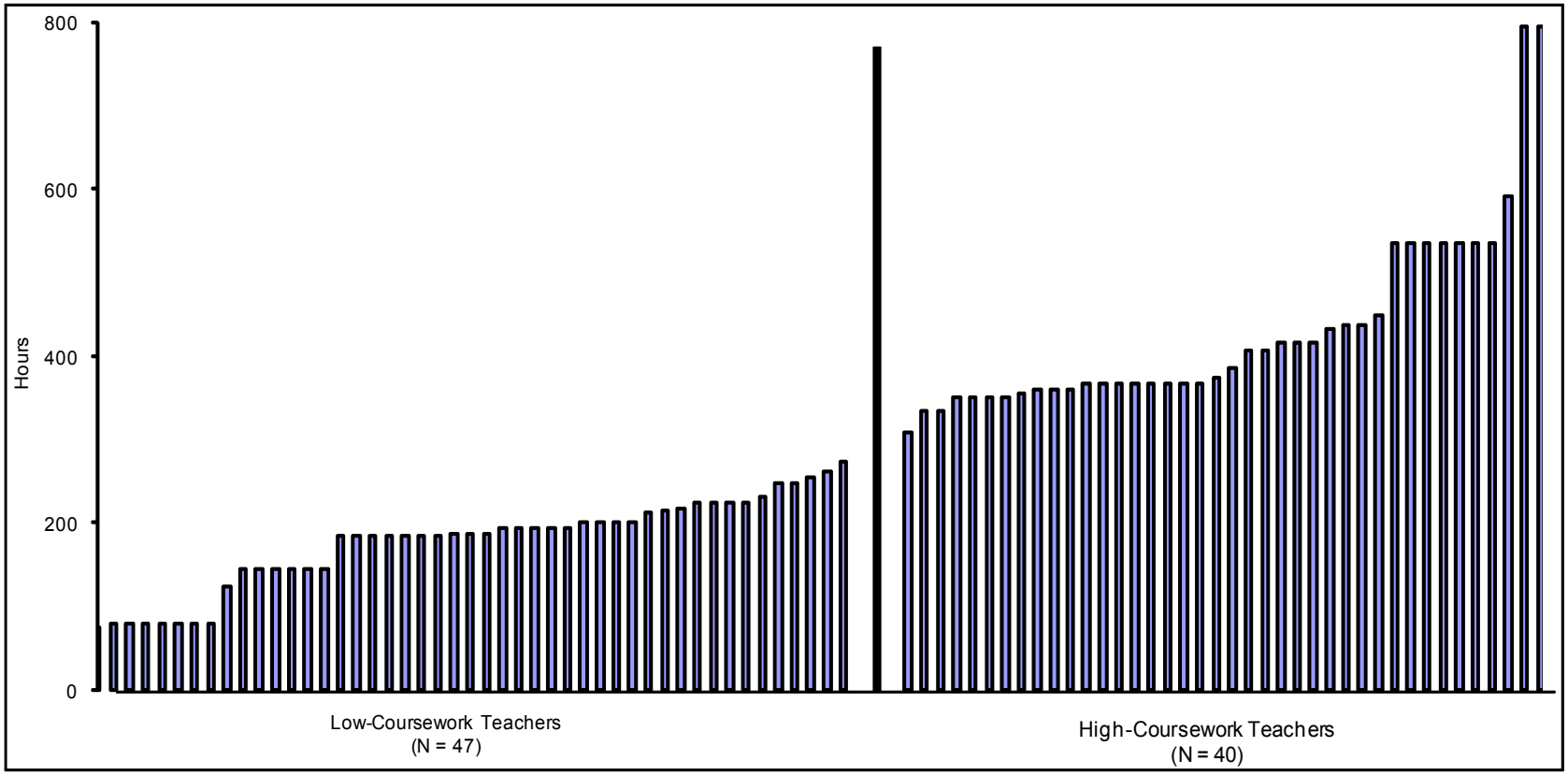
The total amount of instruction required of AC teachers ranged from 75 hours to 795, as shown in Exhibit III.2.[38] If we assume that a typical college course involves 45 hours of instruction (3 hours per week over a 15-week semester), this range represents the equivalent of 1.7 to 17.7 courses. Overall, at the low end of the distribution, about a fourth of the AC teachers (21 of 86) were required to take 185 or fewer hours of instruction, while at the high end, about a fourth (21 of 86) were required to take 375 or more.[39] In the middle, about half the AC teachers (44 of 86) were required to take between 188 and 368 hours of instruction. The mean number of hours across all AC teachers was 296 (SD 153), and the median value was 252.5. The total instruction required of the median teacher was approximately equivalent to the hours of instruction associated with 5.6 college courses.

[37] Three postsecondary institutions in California (California State University, Dominguez Hills; California State University, San Marcos; San Jose State University), two in Louisiana (Northwestern State University; University of Louisiana, Monroe), two in New Jersey (Kean University, Montclair State University), and one in Illinois (Governors State University) also appear on the list of TC program sponsors in Exhibit III.8. Wesleyan College also appears in Exhibit III.8, but we do not count it with the other eight, because it did not operate an AC program independently. We did not systematically explore potential connections or similarities between AC and TC programs operated by the same institution such as the extent to which they required the same courses or the extent to which they shared instructors.

[38] We report unweighted data throughout this chapter. Since the sample for the study is purposive and we collected data on all programs in the study, there is no broader population of programs to which we are drawing inference.

[39] Some AC teachers were still enrolled in their programs and completing coursework during the study. The amount of instruction and fieldwork that we report represents the amount these teachers were expected to complete by the time they finished their programs, within a year or two.

Exhibit III.2. Distribution of Total Hours of Instruction, AC Study Teachers



Source: Interviews with program directors.

Note: This and other exhibits in this chapter exclude one AC teacher who did not enroll in an AC program. However, because she had apparently received no formal instruction toward certification, she is included as part of the low-coursework group in analyses described in later chapters.

To conduct subgroup analyses to examine the effect on teachers' classroom practices and student achievement of having teachers with differing coursework requirements, we divided the AC programs, and thus the teachers that attended the programs, into two groups based on the total instruction their programs required. The main considerations in dividing the sample were (1) creating two roughly equal-sized groups to maximize the precision of the experimental estimates presented in Chapter IV,[40] and (2) selecting a dividing point at which a gap in the distribution of coursework hours would delineate the two groups as clearly as possible. We defined low-coursework AC teachers as those whose program required 274 or fewer hours of instruction, and high-coursework AC teachers as those whose program required 308 or more, as shown in Exhibit III.2. The gap between 274 and 308 was the largest one near the middle of the distribution. Dividing the sample in this way produced a low-coursework group of 46 teachers whose program required, on average, 179 total hours of instruction en route to earning their initial certification (SD 54), equivalent to 3.4 college courses, and a high-coursework group of 40 teachers whose program required, on average, 432 total hours of instruction (SD 112), equivalent to 9.6 college courses.

These analyses are based on the amount of coursework study that teachers were required to complete as part of their training program and not the actual coursework completed during the study year. However, required coursework served as a proxy for actual coursework completed by study teachers, because both TC and AC programs in the study are prescriptive in coursework requirements. For AC programs, there was virtually no room for electives, and all teachers were required to complete similar amounts of coursework. For TC programs, there was more room for variability, but completing all requirements for earning both teacher certification and a bachelor's degree results in few extra courses taken toward the teacher certification.[41] The findings based on the characterization and analyses of high- and low-coursework programs should be interpreted as a program-based requirement, however, and not a finding based on the actual coursework completed by study teachers.

The two AC teacher groups were distributed unevenly across the study states, as shown in Exhibit III.3. Half the high-coursework AC teachers (20 of 40) were in California, about a fourth (11 of 40) were in Texas, and the rest (9) were scattered across three other states. In contrast, a little more than two-thirds (32 of 47) of the low-coursework AC teachers were in Texas, less than one-fifth (9 of 47) were in New Jersey, and the rest (6) were in Louisiana and Wisconsin. The geographic distribution of low- and high-coursework AC teachers reflects state regulations for AC programs, as well as individual program design decisions. For example, California requires that AC programs operated by higher-education institutions provide participants with 36 semester credits of instruction, which we estimate as roughly equivalent to 540 clock hours (assuming 15 clock hours per credit), an amount that ensures that all such programs fall above our low-/high-coursework dividing line. New Jersey, in

[40] By dividing the sample into two subgroups of similar size, we maximize the power of the statistical analysis conducted on each subgroup. Dividing the sample this way allows us to detect effect sizes of 0.20 SD for each subgroup. For more details, see Decker et. al. (2005).

[41] These statements are based on information gathered from the interviews with AC and TC program directors.

Exhibit III.3. Number of Original Low- and High-Coursework AC Teachers, by State

State	Low-Coursework Teachers	High-Coursework Teachers
California	0	20
Georgia	0	4
Illinois, Louisiana, Wisconsin	6	5
New Jersey	9	0
Texas	32	11
Total	**47**	**40**

contrast, requires that AC programs provide participants with approximately 200 hours of instruction, and all the New Jersey programs in the study set their instruction requirements close enough to the state's minimum requirement that they fall below our dividing line. Texas specifies certain topics that AC programs must cover (such as reading instruction), but not the amount of instruction they must provide; the result of this flexibility is that some Texas AC programs have established instructional requirements that fall below our dividing line while others have established requirements that fall above it.[42]

3. Timing of Instruction

AC programs with teachers in the study varied in terms of when they provided instruction to their candidates. We examined the distribution of instruction across three time periods: (1) before candidates became full-time classroom teachers, (2) during the candidates' first year of teaching, and (3) after their first year in the classroom. Ninety percent of the AC teachers (77 of 86) were required to take some instruction from their programs before beginning to teach.[43] Candidates expected to begin in the fall would have received this instruction during the preceding spring or summer. However, nine AC teachers were not required to complete any preliminary coursework; seven of them attended low-coursework programs in New Jersey, and two attended high-coursework programs in California. Ninety-three percent of the AC teachers (80 of 86) were required to take some instruction after starting to teach; the 6 who were required to complete all their coursework before starting were all from low-coursework programs in Texas. Thirty-six percent of the AC teachers (31 of 86) had to complete some coursework after their first year of teaching; this group included 30 high-coursework teachers from programs in California, Georgia, Illinois, Louisiana, and Texas, and 1 low-coursework teacher from a program in New Jersey.

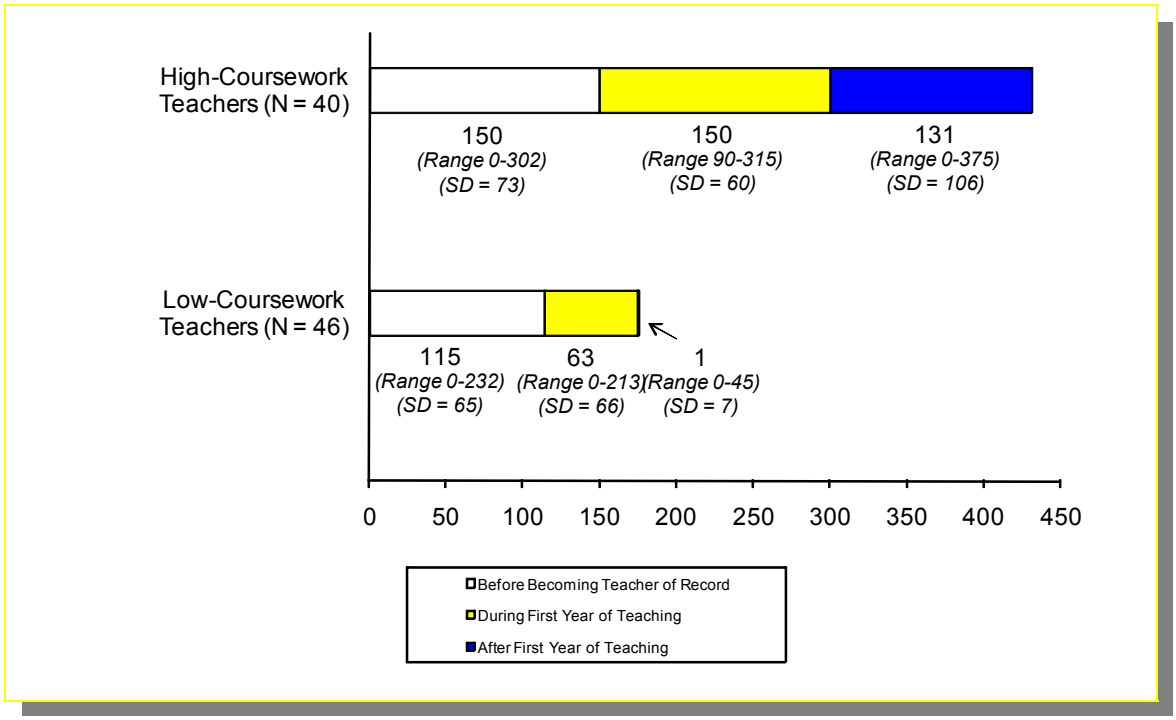
[42] Our sources for state regulations included [www.teach-now.org] and state education department websites, accessed in December 2006 and January 2007.

[43] Percentages and means in the text are rounded to the nearest whole number.

AC teachers' experience also varied in the amount of instruction they were required to take in the three periods. Fifty-nine percent of all AC study teachers (51 of 86)—including 78 percent of low-coursework teachers (36 of 46) and 38 percent of high-coursework teachers (15 of 40)—were required to take at least a plurality of their instructional hours before their full-time teaching assignments began. Nineteen percent overall (16 of 87)—including 21 percent of low-coursework teachers (10 of 47) and 15 percent of high-coursework teachers (6 of 40)—were required to take at least a plurality of their instructional hours during their first year of teaching. And 22 percent overall (18 of 87) were required to take at least a plurality of their instructional hours after their first year; all were high-coursework AC teachers.

On average, high-coursework AC teachers in the study were required to complete more instruction in all three phases—before, during, and after their first year of teaching—than were low-coursework teachers, as shown in Exhibit III.4. Teachers in high-coursework programs were required to take, on average, 150 hours of instruction before they became teachers of record, an additional 150 hours during their first year of teaching, and 131 more hours after their first year. In contrast, low-coursework AC teachers were required to take an average of 115 hours of instruction before they became teachers of record, an additional 63 hours during their first year of teaching, and 1 more hour after their first year. The variation around these means is indicated by the range and SD presented below each mean in Exhibit III.4.

Exhibit III.4. Average Hours of Instruction Relative to First Year of Teaching, Original AC Study Teachers



Source: Program director interviews.

Note: Because of rounding, the high-coursework bar does not sum to the average for total hours reported earlier (432).

The means shown in Exhibit III.4 indicate that high-coursework AC teachers, on average, had greater program-related responsibilities during their first year of teaching than their low-coursework counterparts. The mean of 150 hours for high-coursework AC teachers averages out to 17 hours of instruction per month over a nine-month school year, whereas the mean of 63 for low-coursework teachers averages out to 7 hours per month over a school year, and the difference was statistically significant.

4. Mentoring

To meet "highly qualified" teacher requirements introduced in No Child Left Behind, AC programs or districts hiring AC teachers must provide them with support during at least their first year.[44] We refer to the people who provide these services as "mentors," although their actual titles vary from place to place and from program to program. Eighty-five percent of the AC teachers in our study on whom we have program information (73 of 86) had a program-based mentor assigned to assist them during their first year of teaching after entering the program. The 13 AC teachers who did not have a program-based mentor were from six programs that, according to their directors, did not appoint such mentors, because the new teachers had a mentor appointed by school or district officials.

We did not systematically record the type, amount, frequency, or timing of services provided to AC teachers by program-based mentors. However, the general descriptions we obtained revealed variation across teachers on all these dimensions, either because role expectations differed for mentors from different programs, or because the mentors tailored their services to meet the teachers' specific needs at various points. The AC program directors indicated that program-based mentors provided a range of services as appropriate (for example, assistance with lesson planning, discussing ongoing program coursework, observing the new teachers in action, providing feedback on performance, answering questions, and providing advice or even emotional support). Interactions might be scheduled in advance at various points in the year, such as quarterly classroom observations or weekly check-in discussions, or might take place on an as-needed basis.

B. Comparison of AC Programs in the Study with a Representative Sample of Less Selective Elementary AC Programs in 12 Selected States

Here we compare the programs attended by AC study teachers with the programs in the representative sample in 12 selected states described in Chapter II to determine whether programs included in the study were similar to less selective AC programs in existence in the

[44] Under current federal guidelines, participation in "a program of intensive supervision that consists of structured guidance and regular ongoing support for teachers or in a teacher mentoring program" is one of the criteria that teachers enrolled in AC programs must meet to be considered "highly qualified." (U.S. Department of Education. *Highly Qualified Teachers/Improving Teacher Quality State Grants.* ESEA Title II, Part A, Non-Regulatory Guidance. August 3, 2005.)

12 states at the time study teachers were receiving their training. We examine the following dimensions: admission requirements, coursework requirements, and mentoring.[45]

Similar proportions of programs in the representative sample and programs with teachers in the study require that applicants have a minimum grade point average (GPA), pass a basic skills test, pass a screening interview, submit a writing sample, and pass one or more prerequisite courses; none of the differences between the two groups' means were statistically significant (see Exhibit III.5). Programs with teachers in the study were more likely than those in the representative 12-state sample, at the 0.05 level of statistical significance, to require that applicants submit references or letters of recommendation.

Programs in the study sample and in the representative sample that tied a minimum GPA requirement for admission to all undergraduate coursework set that standard, on average, at about the same level; the mean standard for programs in the representative sample (2.63) was not different at the 0.05 level from the mean standard for programs with teachers in the study (2.57), as shown in Exhibit III.6. Study programs that tied a minimum GPA for admission to a subset of undergraduate courses (such as the last 60 credits) set that standard, on average, higher than did programs in the representative sample with this kind of GPA requirement—2.63 and 2.49, respectively, a difference that was statistically significant.

Exhibit III.5. Types of Admission Requirements Used by Programs in Representative Sample and Programs Attended by AC Teachers in the Study (Percentages)

		AC Programs in Study		
	Programs in Representative Sample (N = 54)	All (N = 43)	Low-Coursework Programs (N = 23)	High-Coursework Programs (N = 20)
Minimum GPA	89[a]	93[b]	91	94[c]
Basic Skills Test	76	77	78	75
Interview	69	58	61	55
Writing Sample	50	58	61	55
References or Letters of Recommendation*	37	65	57	75
Prerequisite Courses	4	7[b]	4	11[c]

Source: Program director interviews.

*Difference between mean for representative sample and mean for all study programs is significant at the 0.05 level.

[a]Number of respondents = 53.

[b]Number of respondents = 41.

[c]Number of respondents = 18.

[45] For this analysis, data on AC study teachers were aggregated up to the program level for teachers who attended the same programs and had the same training experiences.

Exhibit III.6. Average Minimum GPA Requirements for Admission to Programs in Representative Sample and Programs Attended by AC Teachers in the Study

		AC Programs in Study		
	Programs in Representative Sample	All	Low-Coursework Programs	High-Coursework Programs
Minimum GPA for All Undergraduate Coursework	2.63 (n = 38)	2.57 (n = 29)	2.53 (n = 14)	2.60 (n = 15)
Minimum GPA in a Specified Number or Type of Undergraduate Course*	2.49 (n = 21)	2.63 (n = 19)	2.56 (n = 12)	2.76 (n = 7)

Source: Program director interviews.

*Difference between mean for representative sample and mean for all study programs is significant at the 0.05 level.

Programs with teachers in the study required that their candidates take a total amount of instruction, on average, similar to that of programs in the representative sample—303 and 307 hours, respectively, as shown in Exhibit III.7; this difference was not statistically significant. Furthermore, both AC programs in the study and AC programs from the representative sample distributed these hours similarly over time, relative to when their candidates become teachers of record. None of the differences between the groups in hours before, during, or after starting to teach were statistically significant.

Seventy-nine percent of the 43 AC programs with teachers in the study provided a mentor to their participants during their first year of teaching, compared with 70 percent of AC programs in the representative sample.

Overall, these findings indicate that the programs with teachers in the study were similar on 9 of the 11 measured aspects to the programs in the representative sample of less selective AC programs in existence at the time study teachers were receiving their training.

C. CHARACTERISTICS OF TC TEACHERS AND THE PROGRAMS THEY ATTENDED

This section describes the training requirements of TC teachers in the study and the programs they attended. We first identify the sponsors of the TC programs with teachers in the study. Second, we present information on the range in total hours of required coursework that the TC teachers had to complete. Third, we describe required TC teachers' student teaching experiences. Finally, we relate how the TC programs varied on a few structural dimensions.

Exhibit III.7. Average Hours of Instruction Required for Candidates from Programs in Representative Sample and Programs Attended by AC Teachers in the Study

	Programs in Representative Sample (N = 54)	AC Programs in Study		
		All (N = 43)	Low-Coursework Programs (N = 23)	High-Coursework Programs (N = 20)
Before Becoming Teacher of Record	127[a]	124	119	129
During First Year of Teaching	128[b]	117	80	159
After First Year of Teaching	61[c]	66	2	140
Total	**303[d]**	**307**	**201**	**428**

Source: Program director interviews.

[a]Number of respondents = 53.

[b]Number of respondents = 51.

[c]Number of respondents = 50.

[d]Because of missing data, column does not sum to total shown; however, total row includes data for all programs.

1. Sponsoring Institutions

The 87 TC teachers in the study attended programs sponsored by a total of 52 higher-education institutions, as shown in Exhibit III.8. The fact that TC teachers came from more sponsoring institutions than did their AC counterparts (52 versus 28) reflects the process we used to select AC teachers, which focused on particular AC program sponsors that provided large numbers of teachers to specific districts. There was no analogous criterion for their TC program counterparts.

2. Total Hours of Instruction

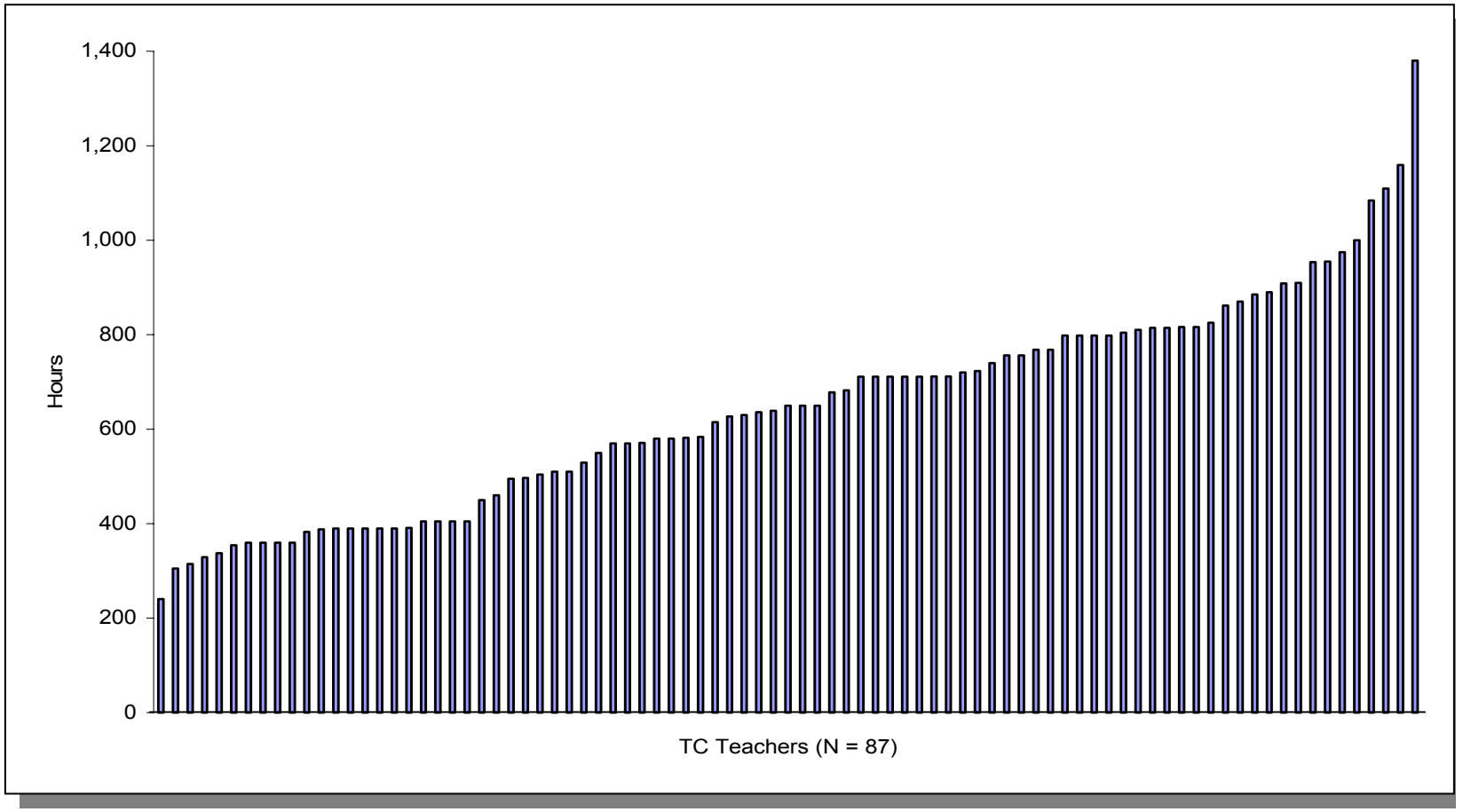
The total amount of instruction required of TC teachers ranged from 240 hours to 1,380 hours, as shown in Exhibit III.9. Assuming that a typical college course involves 45 hours of instruction, this range represents the equivalent of 5.3 to 30.7 courses. At the low end of the distribution, about a fourth of the TC teachers (22 of 86) were required to take 405 or fewer hours of instruction, while at the high end, about a fourth (20 of 86) were required to take 804 or more. In the middle, about half the TC teachers (44 of 86) were required to take 450 to 798 hours of instruction. Overall, TC teachers had to complete an average of 642 hours of instruction (SD 225), equivalent to 14.3 typical college courses; the median value was 644.5 hours.

Exhibit III.8. Sponsors of Programs Attended by Original TC Teachers in Study, by State of Teaching Assignment

State	Sponsor
California	California State University, Dominguez Hills California State University, Long Beach California State University, Northridge California State University, San Marcos Chapman University (California) San Diego State University (California) San Francisco State University (California) San Jose State University (California) University of California, Los Angeles University of San Diego (California)
Georgia	Mercer University (Georgia) Wesleyan College (Georgia)
Illinois	Governors State University (Illinois) Purdue University, Calumet (Indiana)
Louisiana	Louisiana State University, Alexandria Northwestern State University (Louisiana) University of Louisiana, Monroe
New Jersey	Caldwell College (New Jersey) Kean University (New Jersey) Florida A&M University Montclair State University (New Jersey) Ramapo College of New Jersey Rowan University (New Jersey) Rutgers University (New Jersey)
Texas	Abilene Christian University (Texas) Alabama A&M University Angelo State University (Texas) Austin Peay State University (Tennessee) Baylor University (Texas) Dallas Baptist University (Texas) Fisk University (Tennessee) Grambling State University (Louisiana) Ohio State University Sam Houston State University (Texas) Shippensburg University (Pennsylvania) Texas A&M University Texas Christian University Texas Southern University Texas State University Texas Woman's University University of Arkansas, Pine Bluff University of Houston (Texas) University of Houston, Downtown (Texas) University of Libre (Colombia)a University of Mary Hardin-Baylor (Texas) University of North Texas University of St. Thomas (Texas) University of Texas, Arlington University of Texas at Brownsville and Texas Southwest College University of Texas, Pan American University of Wisconsin, Whitewater
Wisconsin	Concordia University (Wisconsin)
Total	**52**

Source: Teacher self-reports on a study eligibility form.

Exhibit III.9. Distribution of Total Hours of Instruction, TC Study Teachers



Source: Interviews with program directors.

To earn certification, TC teachers had to complete, on average, more than twice as much coursework as did AC teachers (642 versus 296 hours, a statistically significant difference), but the two distributions (shown in exhibits III.2 and III.9) overlapped; that is, some AC teachers were required to take more hours of instruction than some TC teachers. This overlap is due partly to state policies that dictate coursework requirements. In some states, such as New Jersey, there was no overlap in required coursework between AC and TC programs within the state. In other states, such as California, the coursework requirements for AC and TC programs were similar; thus the distributions of required coursework were similar for teachers from AC and TC programs.

3. Student Teaching

A standard component of TC programs in general, not just those with teachers in the study, is student teaching. The study TC teachers' experiences ranged from 10 to 21 weeks, typically taking up all or nearly all of a collegiate term (semester or quarter, whichever applied). A common model for student teaching has candidates begin by observing the teacher of record, continue by assuming increasing responsibilities over time, and finish by leading the class several entire days in a row. The study TC teachers were expected to spend, on average, 23 days fully in charge of their classrooms (the range around this mean was 5 to 100 days, and the SD was 19). For 35 of the TC teachers, the expectation was 10 or fewer days, while for 9 it was 45 or more. During their student teaching assignments, TC teachers in the study were visited and observed in action an average of seven times by a program staff member. For each teacher, we asked program directors about how long these sessions lasted on average, including any post-observation debriefing, and found that the average ranged from 28 to 180 minutes. Multiplying the number of observations by the average length of these sessions for each teacher, we found that the TC teachers in the study were observed while student teaching for a total of 10 hours, on average.

4. Variability in TC Program Structure

Whereas AC programs are structurally diverse in terms of the amount of instruction they require before, during, and after their candidates begin teaching, TC programs can be distinguished structurally by the point at which candidates enter the program relative to receiving a bachelor's degree and by the amount of instruction they require. The programs with teachers in this study can be categorized into three broad groups, or models. In one model, undergraduates enter the program near the start of their junior year and take almost exclusively teaching-related courses for two years while earning a bachelor's degree in an education subject (for example, early childhood education or elementary education). In a second model, followed by the New Jersey TC programs in the study, undergraduates enter at about the same point but earn a degree in a field other than education (for example, in the social sciences). These programs are characterized by fewer teaching-related courses, and their requirements are closer to those of a college minor than to those of a major. In a third model, followed by all the California TC programs in the study, as well as one in Ohio and two in Texas, candidates enter the program after completing a bachelor's degree, then take teaching-related courses for 12 or more months.

D. COMPARISON OF AC AND TC TEACHERS' TRAINING EXPERIENCES

1. Instruction and Fieldwork for All Study Teachers

As context for the findings, presented in Chapter IV, on the effects on student achievement and classroom practices experienced by students of teachers from low- and high-coursework AC programs compared to their TC counterparts, Exhibit III.10 presents data on average total hours of instruction, hours in five subject areas of interest (classroom management, reading/language arts pedagogy, math pedagogy, student assessment, and child development), hours in other topics, and hours of fieldwork. The range and SD associated with each mean give a sense of the variability in teachers' training experiences.

Instruction

In this section we present data on four different groups of teachers: (1) teachers who chose low-coursework AC programs, (2) their TC counterparts, (3) teachers who chose high-coursework AC programs, and (4) their TC counterparts. In discussing the average amount of instruction that original study teachers were required to complete as part of their training programs, we examine differences between (1) the low- and high-coursework AC teachers, to explore the extent of differences in their programs' coursework requirements for certification; (2) the two groups of TC teacher counterparts to the low- and high coursework AC teachers, to explore whether they provide a common benchmark for our experimental analyses[46]; and (3) each AC group and its counterpart TC group, to explore differences in coursework requirements that might be related to the results of the experimental and nonexperimental analyses presented in Chapter IV.

Coursework hours data collected for the study focused on five topics: reading/language arts pedagogy, math pedagogy, classroom management, student assessment, and child development. We hypothesized that coursework hours in these specific areas would be most related to student achievement. However, because hours of instruction in topics other than these five accounted for 38 to 51 percent of the average total hours of instruction for each group of teachers, we also discuss hours of such instruction.

Low- and High-Coursework AC Teachers. AC teachers from high-coursework programs were required to take more hours of instruction overall than AC teachers from low-coursework programs, as shown in Exhibit III.10. As discussed above, dividing AC teachers into two similar-sized groups based on a gap in required coursework yielded two groups with large average differences in required coursework. High-coursework AC teachers were required to complete 432 hours of instruction, compared with 179 for low-coursework AC teachers. This difference in total hours of instruction is due to differences in all five subject areas of interest as well as other instruction (defined below). High-coursework AC teachers were required to complete more hours of instruction in all five

[46] If the two groups of TC teachers faced similar instructional requirements in their training programs, then both groups of AC teachers would face similar counterfactuals, and the key analyses (low-coursework AC teachers versus their TC counterparts, and high-coursework AC teachers versus their TC counterparts) would be comparable.

Exhibit III.10. Average Hours of Instruction and Fieldwork, Original Study Teachers

	Low Coursework				High Coursework			
	AC (N = 46)	TC (N = 46)	Difference	*p*-value	AC (N = 40)	TC (N = 40)	Difference	*p*-Value
Instruction								
Classroom management	24 *(Range 8–113) (SD =16)*	54 *(Range 5–183) (SD = 36)*	–30	.00	49 *(Range 8–86) (SD = 19)*	39 *(Range 10–90) (SD = 16)*	11	.01
Reading/language arts pedagogy	26 *(Range 6–60) (SD =21)*	121 (Range 38–361) (SD = 68)	–96	.00	102 *(Range 48–141) (SD = 27)*	109 *(Range 35–195) (SD = 35)*	–7	.31
Math pedagogy	9 *(Range 0–28) (SD = 9)*	41[a] *(Range 0–91) (SD* = 22)	–32	.00	43 *(Range 15–78) (SD = 21)*	41[b] *(Range 5–90) (SD = 19)*	2	.60
Student assessment	16 *(Range 5–43) (SD = 11)*	61 *(Range 10–173) (SD = 32)*	–46	.00	31 *(Range 15–90) (SD = 16)*	55[c] *(Range 6–110) (SD = 25)*	–24	.00
Child development	30 *(Range 5–60) (SD = 20)*	73 *(Range 0–195) (SD = 46)*	–43	.00	41 *(Range 15–100) (SD = 19)*	55 *(Range 5–135) (SD = 32)*	–15	.02
Other	*75 (Range 7–146) (SD = 39)*	321 *(Range 20–818) (SD = 176)*	–247	.00	165 *(Range 28–544) (SD = 110)*	312 *(Range 48–611) (SD = 173)*	–147	.00
Total Instruction	*179[d] (Range 75–274) (SD = 54)*	*671[d] (Range 240–1,380) (SD = 248)*	*–493*	*.00*	*432[d] (Range 308–795) (SD = 112)*	*607[d] (Range 329–975) (SD = 193)*	*–176*	*.00*
Fieldwork	27 *(Range 0–245) (SD = 40)*	*192[a] (Range 0–520) (SD = 108)*	–165	.00	55 *(Range 0–168) (SD = 51)*	153[b] *(Range 0–320) (SD = 112)*	–98	.00

Source: Program director interviews.

[a]Number of respondents = 45.
[b]Number of respondents = 39.
[c]Number of respondents = 38.

subjects, on average, than AC teachers from low-coursework programs: 3.9 times as much instruction in reading/language arts pedagogy, 4.8 times as much in math pedagogy, 2.0 times as much in classroom management, 1.9 times as much in student assessment, and 37 percent more in child development. Although not shown in Exhibit III.10, all these differences were statistically significant at the 0.01 level, except for child development, which was statistically significant at the 0.05 level.

TC Teachers Matched to Low- and High-Coursework AC Teachers. TC teachers matched with low-coursework AC teachers were required to complete a similar amount of total instruction as TC teachers matched to high-coursework AC teachers, 671 hours versus 607, and the difference was not statistically significant. TC teachers matched with low-coursework AC teachers were required to complete the same amount or more instruction in each of the five subject areas, on average, than were TC teachers matched with high-coursework AC teachers, with statistically significant differences for classroom management and child development (at the 0.05 level; analysis not shown in Exhibit III.10). Thus, in terms of required coursework, TC teachers matched to low- and high-coursework AC teachers served as a common benchmark in conducting the subgroup analysis.

Matched AC and TC Teacher Subgroups. AC teachers from low-coursework programs were required to complete, on average, about one-quarter of the total hours of instruction overall as their TC counterparts (179 hours versus 671 hours). In addition, they were required to complete less coursework in all subject areas of interest. For example, their programs required about one-fifth the instruction in reading/language arts pedagogy (26 versus 121 hours), less than one-fourth in math pedagogy (9 versus 41 hours), and less than half in classroom management (24 versus 54 hours). All the differences were statistically significant.

AC teachers from high-coursework programs were required to complete, on average, less instruction than their TC counterparts, 432 hours versus 607 hours, a difference that was statistically significant. They were required to complete less coursework in two topics of interest (student assessment and child development), with the differences statistically significant. However, their programs required *more* instruction in classroom management (49 versus 39 hours), a difference that was statistically significant. There was no statistically significant difference in the amount of math pedagogy instruction (43 versus 41). Considering all five topics of interest together (that is, excluding "other" instruction), high-coursework AC teachers' programs required 91 percent as much instruction as their TC counterparts' programs (267 versus 295 hours), a difference that was statistically significant at the 0.05 level.

The distribution of the differences in required coursework between each AC teacher and his or her TC counterpart was also large. Exhibit III.11 shows the difference between each low-coursework AC teacher and the TC counterpart. The average difference in required coursework between low-coursework AC teachers and their TC counterparts was driven by two factors: (1) each low-coursework AC teacher was required to complete fewer hours of coursework than the TC counterpart, and (2) in more than half the pairs (30 of 48) this difference was 400 hours or more, equivalent to 8.9 or more courses. For high-coursework AC teachers, more than half (27 of 40) were also required to complete fewer hours of instruction than their TC counterparts, but the difference was 400 or more hours in only 11 of 40 of the pairs. In addition, 15 of 40 low-coursework AC teachers were required to complete the same or more hours of coursework as their TC counterparts.

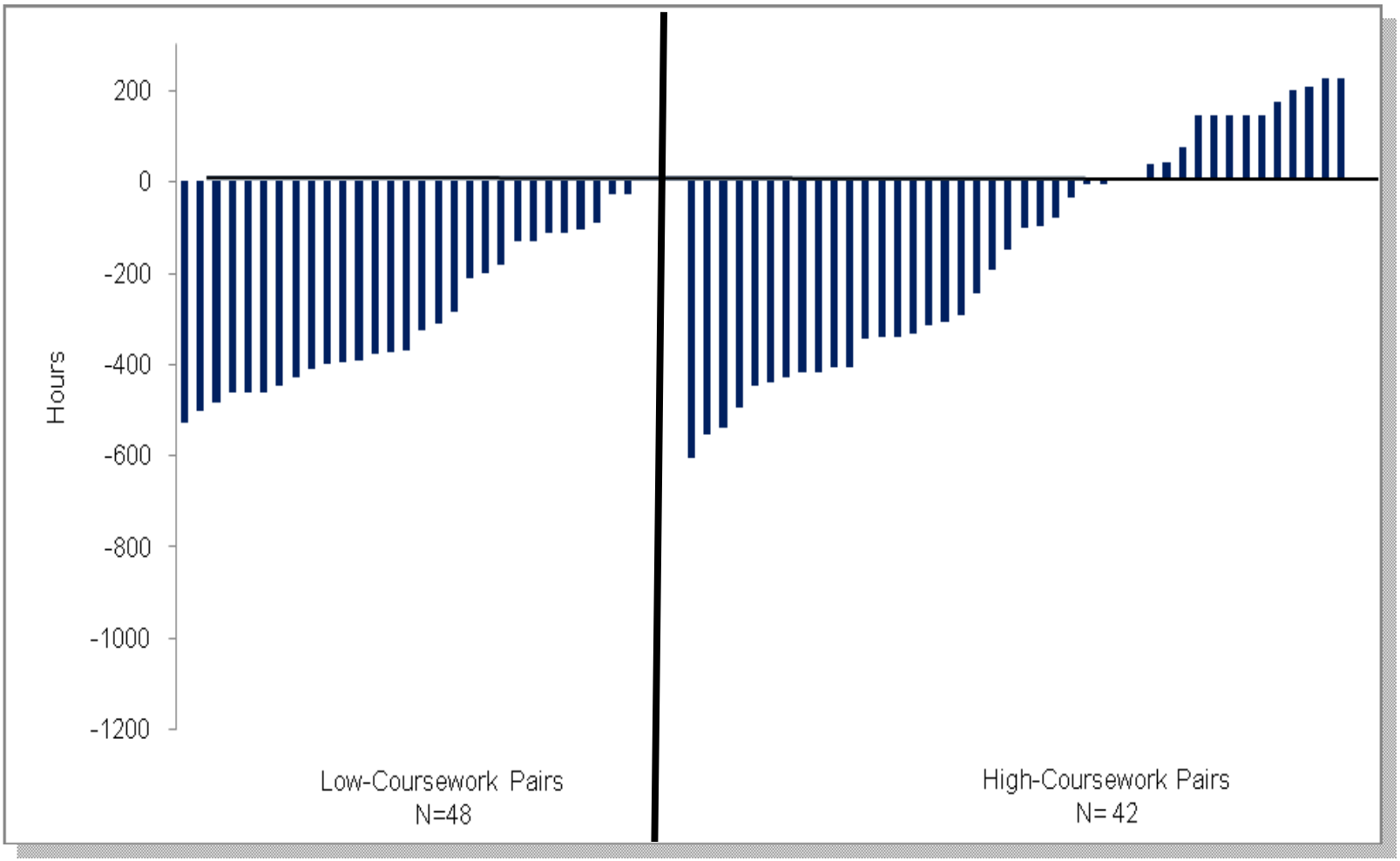
"Other" Instruction. For all teachers, some of the required coursework fell outside the five subjects of most interest in this study. Instruction in other topics accounted for, on average, 42 percent of total coursework for the low-coursework AC teachers, 48 percent for their TC counterparts, 38 percent for the high-coursework AC teachers, and 51 percent for their TC counterparts. "Other" instruction accounted for half the statistically significant 493-hour difference in total instruction between low-coursework AC teachers and their TC counterparts, and for 84 percent of the statistically significant 176-hour difference between high-coursework AC teachers and their TC counterparts.[47]

Fieldwork

The average hours of fieldwork required for each of the four teacher groups is shown in Exhibit III.10. High-coursework AC teachers were required to conduct about twice as many hours of fieldwork, on average, as low-coursework AC teachers (55 versus 27 hours, a statistically significant difference). But AC teachers from both low- and high-coursework

[47] Because of the way program information was collected, data are not available to describe fully the content of "other" instruction for any of the four teacher groups. As explained in Chapter II, "other" consists of any hours of instruction, whether a whole course or part of a course, that were not counted toward one of the five areas of interest in this study. A complete accounting of "other" therefore depends on knowing the number of instructional hours that each required course contributed to each of the five areas of interest. In obtaining estimates from program directors of the hours of instruction provided in each of the five areas of interest, however, we did not require them to document how many hours in each area came from particular courses on the program's full list of required courses. Therefore, we do not always know whether a given course contributed all, some, or none of its instructional hours to the "other" category. Courses that likely were counted entirely or largely toward "other" instruction included methods courses focused on subjects other than reading and math (for example science, social studies, art, music, health, physical education), general methods courses (for example, cooperative learning, ability grouping), courses on using technology, courses on educational psychology (for example, learning theories), courses on dealing with special needs students (students with disabilities, gifted students), courses on working with students from diverse cultural backgrounds, courses on social issues or trends (for example, "Democratic Society," "Equity and Justice"), courses on gang or drug awareness, courses on dealing with families and communities (for example, parent involvement), foundations courses (for example, "Foundations of Education," "Cornerstones of the Profession"), courses on legal issues, and courses on safety (for example, first aid, crisis prevention).

Exhibit III.11 Distribution of Differences in Required Coursework Between Each AC Teacher and Their TC Counterpart



Source: Interviews with program directors.

programs were required to conduct less fieldwork, on average, than their TC counterparts. High-coursework AC teachers conducted, on average, one-third as much fieldwork as their TC counterparts (55 versus 153 hours), and low-coursework AC teachers were required to conduct, on average, less than one-fifth as much (27 versus 192 hours); both differences were statistically significant.

Summary

These analyses of coursework and fieldwork indicate that the coursework and training required for study teachers varied in ways consistent with the study design. The division of AC teachers into those who attended low- and high-coursework programs yielded distinct AC program subgroups with, on average, different amounts of required coursework. The amount of instruction received by their counterpart TC teachers, however, was not different, which indicates that TC teachers to whom both low- and high-coursework AC teachers were compared were required to complete a similar amount of total instruction. Low-coursework AC teachers were required to take only one-third the hours of instruction of their TC counterparts, which supports a test of the effect on student achievement and instructional practices experienced by students of AC teachers who were required to complete relatively little coursework. High-coursework AC teachers and their TC counterparts were required to take similar amounts of instruction across the five topics of interest, but fewer hours of instruction in other areas, which also supports a test of the effect of AC teachers, but with a smaller difference in total hours of required coursework. Comparisons of both low- and high-coursework AC teachers to their TC counterparts provides a test of the effect of the timing of coursework requirements, since all AC teachers were allowed to begin teaching in the classroom before they completed their coursework requirements.

2. Variable Experiences Across and Within States

As context for state-level impact analyses (presented as additional subgroup analyses in Chapter IV), we now provide some examples of how study sample teachers' experiences with required coursework varied across states, due to variability in state policies and institutional practices regarding AC and TC programs. The 20 AC teachers in California were all from high-coursework programs, as we defined them, and were required to take an average of 488 hours of instruction, whereas the 18 TC teachers in California were required to take an average of 448 hours, and both AC and TC groups began their training after receiving bachelor's degrees. This represents the smallest difference in required coursework between AC and TC teachers among all the states in the study. In contrast, the eight New Jersey AC teachers on whom we have program data all earned a bachelor's degree before beginning their training and attended low-coursework programs that required an average of 200 hours of instruction, whereas eight TC teachers trained in New Jersey completed their training as undergraduates and were required to take an average of 394 hours.[48] Finally, although one-fourth of the AC teachers in Texas (11 of 43) were in high-coursework programs, the average instruction required for Texas AC teachers was 216 hours, below the

[48] A ninth TC teacher in New Jersey was trained in another state and was required to complete 955 hours of coursework while an undergraduate.

full AC sample average of 294, whereas their TC counterparts—of whom 77 percent (34 of 44) attended undergraduate-based in-state programs—received an average of 756 hours of instruction, above the full TC sample average of 642. This represents the largest difference in required coursework between AC and TC study teachers among all the states.

E. Comparison of AC and TC Teachers' Background Characteristics and Professional Experiences

This section examines the background characteristics and professional experiences of teachers in the study, using primarily data from the teacher survey.[49] Differences in students' test scores or teachers' classroom practices for teachers who chose to attend different types of programs could reflect preexisting personal differences rather than the training they received. There is some evidence, for example, that minority students benefit from having a teacher of the same race (Dee 2004; Clotfelter et al. 2007). Teachers with different levels of cognitive ability may demonstrate different effects on student achievement, regardless of their teacher preparation route (Ferguson and Ladd 1996). And years of experience may matter, as research has shown that teachers' classroom performance improves between their first and their second or third year of teaching but then stabilizes (Boyd et al. 2005; Hanushek et al. 2005).

1. Background Characteristics

Demographics. AC teachers differed from TC teachers on a few demographic characteristics, as shown in Exhibit III.12. Both groups of AC teachers were more likely than their TC counterparts to identify themselves as black (40.5 percent versus 17.5 percent and 32.4 percent versus 7.5 percent) and less likely to identify themselves as white (50 percent versus 75.5 percent and 40.5 percent versus 70 percent), and all differences were statistically significant. AC and TC teacher groups did not differ, however, on other measures of race/ethnicity. Both groups of AC teachers were older than their TC counterparts, on average (33.5 years versus 28.3 years and 33.9 years versus 30.1 years), and the difference was statistically significant. Finally, low-coursework AC teachers were more likely than their TC counterparts to have children (70.2 percent versus 28.3 percent), and this difference was also statistically significant.

Educational Attainment and Cognitive Ability. Both low- and high-coursework AC teachers were less likely than their TC counterparts to report having majored in education as undergraduates (2.2 percent versus 78.3 percent and 21.4 percent versus 56.8 percent), and the differences were statistically significant, as shown in Exhibit III.13. High-coursework AC teachers were more likely than their TC counterparts to be taking courses toward an advanced degree or teaching certification (57.1% versus 29.5%), and the difference was statistically significant. The difference between low-coursework AC teachers and their TC counterparts on this variable was not statistically significant. On two possible indicators of

[49] In most of the subsequent exhibits in this chapter, the maximum number of respondents in each of the four teacher groups is greater than in the earlier exhibits, because teachers who were in the study both years are included twice if they provided two years of data. Some of the characteristics of these teachers, such as race and undergraduate major, do not change across the two years, but some do change, such as whether they were currently taking coursework.

the cognitive ability of teachers before they received any training to become teachers—college entrance exam scores and selectivity of undergraduate institution—there were no statistically significant differences between either group of AC teachers and their TC counterparts.

Exhibit III.12. Teacher Demographics (Percentages, Except Where Noted)

	Low-Coursework AC Teachers and Their TC Counterparts				High-Coursework AC Teachers and Their TC Counterparts			
	AC	TC	Difference	*p*-Value	AC	TC	Difference	*p*-Value
Race/Ethnicity[a]								
White	50.0	75.5	–25.5	0.02	40.5	70.0	–29.5	0.01
Black	40.5	17.5	23.0	0.03	32.4	7.5	24.9	0.01
Hispanic/Latino	15.5	15.2	0.3	0.97	25.0	16.3	8.7	0.34
Other[b]	0.0	0.0	0.0	0.33	8.1	12.5	–4.4	0.54
Female	95.7	97.8	–2.2	0.56	78.6	88.6	–10.1	0.22
Have Children	70.2	28.3	43.5	0.00	38.1	29.5	8.5	0.41
Average Age (Years)	33.5	28.1	5.5	0.00	33.9	30.1	3.8	0.01
Sample Size (Range)	**42–46**	**40–46**			**37–42**	**40–44**		

Source: Teacher survey.

[a]Categories were not mutually exclusive.

[b]Combines three original response categories: Asian, Native Hawaiian or Pacific Islander, and American Indian or Alaska Native.

Exhibit III.13. Teacher Education and Cognitive Ability (Percentages, Except Where Noted)

	Low-Coursework AC Teachers and Their TC Counterparts				High-Coursework AC Teachers and Their TC Counterparts			
	AC	TC	Difference	*p*-Value	AC	TC	Difference	*p*-Value
Education major	*	78.3	*	0.00	21.4	56.8	–35.4	0.00
Highest degree: bachelor's	82.6	91.3	–8.7	0.22	76.2	77.3	–1.1	0.91
Highest degree: master's	17.4	8.7	8.7	0.22	23.8	22.7	1.1	0.91
Currently taking courses[a]	30.4	19.6	10.9	0.24	57.1	29.5	27.6	0.01
Selective undergraduate institution[b]	15.0	31.0	–16.0	0.09	26.3	33.3	–7.0	0.50
Average SAT or equivalent composite score[c] (points)	923	959	–35.8	0.33	1,010	1,013	–2.5	0.96
Sample Size	**46**	**46**			**42**	**44**		

Sources: Teacher survey for all but SAT scores, which were obtained from the College Board, and ACT scores, which were obtained from ACT.

* - Values suppressed to protect respondent confidentiality.

[a]Includes courses toward an advanced degree or teaching certification.

[b]Sample sizes for this item were 40 for low-coursework AC teachers, 42 for their TC counterparts, 38 for high-coursework AC teachers, and 42 for their TC counterparts.

[c]We converted ACT scores to SAT equivalents using the concordance procedure available from the College Board. Sample sizes for this item were 38 for low-coursework AC teachers, 40 for their TC counterparts, 28 for high-coursework AC teachers, and 32 for their TC counterparts.

2. Professional Experiences

Teaching and Other Classroom Experience. As shown in Exhibit III.14, low-coursework AC teachers reported, on average, 0.7 fewer years of full-time teaching experience than their TC counterparts, a difference that was statistically significant. High-coursework AC teachers reported, on average, 0.4 more years of experience as emergency certified teachers than their TC counterparts; this difference was also statistically significant.

Mentoring and Other Support. Both AC teacher subgroups were more likely than their TC counterparts to report having worked with a mentor, master teacher, or field supervisor (hereafter, "mentor") in their first year of teaching (93.5 percent versus 78.3 percent and 90.5 percent versus 65.9 percent), and both differences were statistically significant, as shown in Exhibit III.15[50] Both subgroups of AC teachers were also more likely than their TC counterparts to report having had a second mentor during their first year (48.8 percent versus 2.8 percent and 36.8 percent versus 13.8 percent). In addition, high-coursework AC teachers were more likely than their TC counterparts to report having had opportunities during their first year of teaching to observe other teachers' classrooms (90.5 percent versus 72.7 percent). All these differences related to support were statistically significant at the 0.05 level.

Exhibit III.14. Average Years of Teaching and Other Classroom Experience, Including First Year in Study[a]

	Low-Coursework AC Teachers and Their TC Counterparts				High-Coursework AC Teachers and Their TC Counterparts			
	AC	TC	Difference	*p*-Value	AC	TC	Difference	*p*-Value
Full-time Teaching								
Certified teacher	2.4	3.0	–0.6	0.06	2.7	2.8	–0.2	0.51
Emergency certified teacher	0.2	0.3	0.0	0.75	0.6	0.2	0.4	0.03
Long-term substitute	0.1	0.2	–0.1	0.20	0.1	0.2	–0.1	0.43
Subtotal	*2.8*	*3.5*	*–0.7*	*0.04*	*3.5*	*3.2*	*0.2*	*0.45*
Other Experience								
Teacher aide	0.4	0.3	0.2	0.56	1.0	0.6	0.4	0.36
Short-term substitute	0.5	0.6	-0.1	0.77	0.6	0.6	0.1	0.78
Other position	0.2	0.0	0.2	0.28	0.1	0.1	0.0	0.85
Sample Size	**46**	**46**			**42**	**44**		

Source: Teacher survey.

[a]Teachers with some experience in any given category, but less than a year, were instructed to round up to one year.

[50] As discussed in Section A of this chapter, based on AC program director interviews, 85 percent of the AC teachers in the study had a program-based mentor in their first year of teaching after entering the program, and the rest were reported by program operators to have had a school- or district-based mentor. The fact that less than 100 percent of AC survey respondents reported having a mentor could reflect measurement error in the survey or the program director interviews (for example, faulty memories regarding mentoring, differing interpretations of our questions or definitions, or inaccurate assumptions about what support would be provided by districts or schools).

Exhibit III.15. Mentoring and Support During First Year of Teaching (Percentages)

	Low-Coursework AC Teachers and Their TC Counterparts				High-Coursework AC Teachers and Their TC Counterparts			
	AC	TC	Difference	*p*-Value	AC	TC	Difference	*p*-Value
Had a mentor	93.5	78.3	15.2	0.04	90.5	65.9	24.6	0.01
Had a second mentor[a]	48.8	*	*	0.00	36.8	13.8	23.0	0.04
Seminars or classes for beginning teachers	84.4	76.1	8.4	0.33	88.1	81.8	6.3	0.43
Extra classroom assistance (e.g., teacher aide, team teaching)	31.1	41.3	–10.2	0.32	38.1	29.5	8.5	0.41
Regular supportive communication with school officials	80.0	67.4	12.6	0.18	71.4	52.3	19.2	0.07
Opportunities to observe other teachers	77.6	58.7	19.1	0.06	90.5	72.7	17.7	0.04
Sample Size	**45**[b]	**46**			**42**	**44**		

Source: Teacher survey.

* - Values suppressed to protect respondent confidentiality.

[a]Number of respondents = 43 for low-coursework AC teachers, 36 for their TC counterparts, 38 for high-coursework AC teachers, and 29 of their TC counterparts.

[b]Number of respondents = 46 for item on first mentor.

Teachers with at least one mentor in their first year were asked about the frequency of various interactions with their mentors. Both AC groups reported higher frequencies than their TC counterparts on all items, and all measured differences were statistically significant, as shown in Exhibit III.16. Teachers who had at least one formal meeting with their mentor were asked about the average length of these meetings. High-coursework AC teachers reported longer average formal meetings with their first mentor compared to their TC counterparts (36.4 minutes versus 24.1 minutes), and the difference was statistically significant, whereas the difference between low-coursework AC teachers and their TC counterparts was not statistically significant.

Professional Development. Teachers were asked whether any of eight specific topics had been covered in school- or district-supported professional development they had received during their first three years of teaching. As shown in Exhibit III.17, a higher percentage of both low- and high-coursework AC teachers received professional development in methods of teaching/pedagogy than their TC counterparts, and both differences were statistically significant. High-coursework AC teachers were also more likely than their TC counterparts, by a statistically significant margin, to have received professional development in student discipline and classroom management. A statistically significant higher percentage of high-coursework AC teachers reported spending 11 or more days in professional development during their first year of teaching than did their TC counterparts. The difference between low-coursework AC teachers and their TC counterparts was not

statistically significant. Neither AC group differed from its counterpart TC group by a statistically significant margin in their second or third year of teaching.

Exhibit III.16. Frequency of Mentoring Activities in First Year of Teaching[a]

	Low-Coursework AC Teachers and Their TC Counterparts				High-Coursework AC Teachers and Their TC Counterparts			
	AC	TC	Difference	*p*-Value	AC	TC	Difference	*p*-Value
Mentor observed classroom teaching	19.7	7.0	12.7	0.00	17.7	8.9	8.9	0.01
Teacher observed mentor's teaching	10.7	4.9	5.8	0.03	8.4	4.2	4.2	0.05
Received written feedback from mentor	22.0	6.0	16.1	0.00	17.9	7.8	10.2	0.00
Met formally with mentor	28.6	14.3	14.4	0.00	22.7	11.3	11.3	0.01
Met informally with mentor	29.2	18.5	10.8	0.01	26.0	15.5	10.5	0.01
Average length of formal meetings with first mentor (minutes)[b]	26.6	20.2	6.5	0.07	36.4	24.1	12.3	0.01
Sample Size[c]	**46[d]**	**46**			**42[e]**	**44**		

Source: Teacher survey.

[a]Except where noted, the table presents the average number of times each activity occurred during the first year of teaching for those teachers who had at least one mentor, although means include frequency of activities across two mentors, for teachers who had a second mentor. Response categories included "never," which we coded as 0; "one time only," which we coded as 1 time in a year; "2-3 times a term," which we coded as 8 times in year; "at least once a month," which we coded as 10 times in a year; and "at least once a week," which we coded as 36 times in a year.

[b]Question posed only to teachers who met formally with their first mentor. The number of respondents upon which means were calculated was 40 low-coursework AC teachers, 32 of their TC counterparts, 35 high-coursework AC teachers, and 28 of their TC counterparts. Response categories included "15 minutes or less," which we coded as 7.5 minutes; "15 to 30 minutes," which we coded as 22.5 minutes; "30 to 60 minutes," which we coded as 45 minutes; and "more than 60 minutes," which we coded as 90 minutes. Teachers whose meetings lasted an average of 30 minutes would have had to choose either "15 to 30" or "30 to 60."

[c]With one exception, described above, sample sizes in this table reflect the maximum number of respondents for each group of teachers. Teachers who did not have a mentor and would have skipped these items were given a value of 0 on these items so that the results would convey frequency of mentoring activities across all study teachers.

[d]Number of respondents = 45 for item on frequency of written feedback.

[e]Number of respondents = 41 for item on frequency of informal meetings.

Exhibit III.17. Content and Amount of Professional Development (Percentages)

	Low-Coursework AC Teachers and Their TC Counterparts				High-Coursework AC Teachers and Their TC Counterparts			
	AC	TC	Difference	*p*-Value	AC	TC	Difference	*p*-Value
Content Areas Covered in First Three Years								
Standards (content and performance) in an area taught	84.8	89.1	–4.3	0.55	85.7	93.2	–7.5	0.27
Methods of teaching/pedagogy	80.4	60.9	19.6	0.04	82.9	63.6	19.3	0.05
Selecting exemplary instructional materials	39.1	45.7	–6.5	0.54	48.8	50.0	–1.2	0.92
Applications of technology to instruction	69.6	65.2	4.3	0.67	57.1	50.0	7.1	0.52
Student assessment	76.1	69.6	6.5	0.49	82.9	70.5	12.5	0.18
Student discipline and classroom management	69.6	69.6	0.0	1.00	73.2	38.6	34.5	0.01
Study of reading/ language arts	84.8	84.8	0.0	1.00	90.5	84.1	6.4	0.39
Study of math	69.6	82.6	–13.0	0.15	81.0	75.0	6.0	0.52
Amount								
11 or more days in professional development in first year	58.7	50.1	8.7	0.41	61.9	38.6	23.3	0.04
11 or more days in professional development in second year[a]	50.0	31.0	19.0	0.08	52.6	41.0	11.6	0.32
11 or more days in professional development in third year[b]	39.1	33.3	5.8	0.67	53.3	50.0	3.3	0.82
Sample Size	**46**	**46**			**42**[c]	**44**		

Source: Teacher survey.

[a]Item not applicable to teachers who had taught for only one year. Number of respondents upon which means were calculated: 40 for low-coursework AC teachers, 42 for their TC counterparts, 38 for high-coursework AC teachers, 39 for their TC counterparts.

[b]Item not applicable to teachers who had taught for only one or two years. Number of respondents upon which means were calculated: 23 for low-coursework AC teachers, 30 for their TC counterparts, 30 for high-coursework AC teachers, 24 for their TC counterparts.

[c]Number of respondents = 41 for items on methods of teaching/pedagogy, selecting exemplary instructional materials, student assessment, and student discipline and classroom management.

F. SUMMARY

This chapter presented information on a sample of less selective AC elementary teacher preparation programs and a set of teachers who attended such programs in recent years, as well as on a set of relative novice TC teachers who taught in the same grade levels at the same schools. Key findings include:

- **Both the AC and the TC programs with teachers in the study were diverse in the total instruction they required for their candidates.** The total hours

required by AC programs ranged from 75 to 795, and by TC programs, from 240 to 1,380. Thus not all AC programs require fewer hours of coursework than all TC programs. One-fifth of the AC teachers in this study were required to take as much or more instruction than one-fourth of the TC teachers. The degree of overlap was dictated by variations in state policies on teacher certification programs. For example, in New Jersey all AC teachers were required to complete fewer hours of coursework than all TC teachers, while in California, the range of coursework hours required was similar for AC and TC teachers.

- **While teachers trained in TC programs receive all their instruction (and participate in student teaching) prior to becoming regular full-time teachers, AC teachers do not necessarily begin teaching without having received any formal instruction.** Overall, low-coursework AC teachers in the study were required to take an average of 115 hours of instruction—64 percent of the total amount of instruction they would receive—before starting to teach, and high-coursework AC teachers in the study were required to take an average of 150 hours—about 35 percent of the total amount they would receive—before starting to teach. Nine AC teachers in the study, seven of them from New Jersey, were not required to complete any coursework before becoming regular full-time teachers.

- **On most topics for which we measured hours of instruction, low- and high-coursework AC teachers were required to complete less coursework, on average, than their TC counterparts.** Every difference we examined between low-coursework AC teachers and their TC counterparts was statistically significant at the 0.01 level; the former were required to take an average of 493 hours less total instruction than the latter. Although the AC and TC teachers in this study were required to take different amounts of instruction, on average, both overall and in certain topic areas, the two kinds of programs devoted a similar proportion of their total instructional time to certain topics. Instruction required for AC teachers was not more focused on the core subjects of reading and math pedagogy, for example, than instruction was for TC teachers. The AC teachers' programs devoted an average of 27 percent of their instructional time to these two subjects combined; among TC teachers' programs it was 26 percent, which was not a statistically significant difference.

- **Over 92 percent of the AC teachers in this study were reported to have had a program- or school-based mentor during their first year of teaching; in contrast, about three-fourths of the TC teachers reported having had a mentor in their first year.** Although AC teachers may begin teaching with fewer hours of instruction and less firsthand exposure to elementary classrooms than TC teachers, their programs and the schools and districts are more likely to provide direct support to help AC teachers adapt to their new responsibilities.

- **There were no statistically significant differences between the AC and TC teachers in this study in their average scores on college entrance exams, the selectivity of the college that awarded their bachelor's degree, or their**

level of educational attainment. Both low- and high-coursework AC teachers were more likely than their TC counterparts to identify themselves as black (40.5 percent versus 17.5 percent and 32.4 percent versus 7.5 percent) and less likely as white (50 percent versus 75.5 percent and 40.5 percent versus 70 percent). In addition, the low-coursework AC teachers were more likely than their TC counterparts to report having children (70.2 percent versus 28.3 percent).

CHAPTER IV

ANALYSES AND FINDINGS

This study seeks to inform two distinct policy questions: (1) What are the relative effects on student achievement of teachers who chose to be trained through different routes to certification, and how do observed teacher practices vary by chosen route to certification? and (2) What aspects of certification programs (for example, amount of coursework, timing of coursework relative to being the lead teacher in the classroom, core coursework content) are associated with the teacher effectiveness?

The empirical evaluation provides information to help answer these two questions. For the first, we rely on experimental methods that measure the differences in test scores of students who were randomly assigned to either AC or TC teachers, as well as differences in teacher classroom practices. To address the second question, we rely on nonexperimental methods to estimate the relationship between student outcomes and teacher training. Because we cannot experimentally separate the characteristics of the teacher from those of the program the teacher chose, our nonexperimental estimates are suggestive of program and teacher characteristics that may be associated with differences in teacher effectiveness, and cannot be interpreted causally.

A. EXPERIMENTAL ANALYSES

Schools in the study had at least one AC and one TC teacher in the same grade level to whom students were randomly assigned. This created teacher pairs that could facilitate, across schools and grade levels, a series of mini-experiments to examine differences in test scores in reading and math. The purpose of within-school random assignment was to minimize preexisting differences, in students and schools, that might contribute to subsequent differences in average test scores. Randomization of students equates, on average, the classroom characteristics taught by each pair, and school differences are eliminated since each mini-experiment takes place in the same school and grade. For example, in a given experiment, the experiment-level effect on reading scores provides an unbiased estimate of the teachers' effect on the achievement of their students as measured by the difference between the average reading test scores of students assigned to the AC

teacher versus the TC teacher.[51] To calculate the overall AC effect for the low- and high-coursework teachers, we took the simple average of the experiment-level effects for each group of AC teachers.[52,53] As discussed in Chapter II, this estimation strategy shows the effect on student achievement of AC teachers compared to their TC counterparts. Therefore, the estimates represent the differences in student outcomes that would be expected if an AC teacher instead of a TC teacher were placed in a classroom in the study schools.[54] Because the effect is generated by a combination of teachers' pretraining characteristics, their training, and school hiring practices, it does not show the relative effect of AC programs.

All students in the sample were tested in mathematics and reading. As our primary student outcome, we used the normal curve equivalent (NCE) of each student's test as a measure of the student's reading and mathematics ability at the end of the intervention. A simple comparison of the posttest NCE scores would provide an unbiased estimate of the impact of the teachers. However, to improve the precision of our estimates, we regression-adjust the posttest means to control for other student characteristics that can affect posttest performance, namely, the student's scores on the tests at the beginning of the year, gender, race/ethnicity, and eligibility for free or reduced-priced lunch. The regression also controls for the experience of the teacher. Throughout the study, the regression-adjusted means are presented. The means for the AC classes are equal to the unadjusted TC mean plus the adjusted difference. Full details of the estimation, including unadjusted posttest means, are presented in the Appendix.

This experimental study assigned students to either AC or TC teachers. Because the students were the unit of random assignment, the measured effects are interpreted as the influence on the students from being placed in the classroom of an AC or a TC teacher. Throughout the chapter, we use the shorthand terms "teacher effects" and "the effectiveness of teachers" to indicate the effect on student achievement or teaching practices experienced by a student's being placed in the classroom of an AC teacher.

---

[51] The estimates control for the experience of the teacher, because this varies between AC and TC teachers in our sample.

[52] See Appendix A for full details of the estimation strategy.

[53] Since 14 teachers were in the study for both years, there is potential for a slight dependence effect due to a repeated teacher effect. It is minimized, however, by having new students assigned to the teacher each year and in 4 cases, a new comparison teacher for the teacher in the study both years. Therefore, we did not attempt to adjust for this effect in the analyses.

[54] The estimates in all the experimental analyses include all study teachers, regardless of whether a teacher moved during the year. When the original study teacher left the classroom during the year, we obtained the "intent-to-treat" estimates by averaging the effects on test scores according to the treatment status of the original teacher. The study focuses on the intent-to-treat estimates because they best answer what would be experienced by students who are taught by an AC or a TC teacher, which includes the probability that a teacher will leave.

1. Student Test Scores

There was no statistically significant mean difference in the test scores of students taught by AC teachers and the scores of students of their TC counterparts. This finding was robust across subgroups (that is, teachers from AC programs with low- or high-coursework requirements) to the grade level of the students and to different measures of teacher experience. There was some evidence, however, of heterogeneity of the effects across the states in the study.

Though the mean effects were not statistically different from zero, the effects across mini-experiments ranged size from –1 to 0.9. However, because of the small sample sizes of individual classrooms, and because only one teacher pair is represented at each of the extreme values, these effects should be interpreted cautiously.[55] The extreme values on either side of the individual effect size distribution represent a range of –1.1 to 1.1 grade levels of learning.[56] In other words, in at least one case the students in the AC teacher's class measured more than a full grade level below the students of the TC counterpart, and in at least one case the AC teacher's class was more than a full grade level above the TC counterpart's. Therefore, the mean effect size of all the experiments masks information about the effect on student achievement of a particular AC teacher compared to the TC counterpart. Thus, it may be very difficult to predict, based solely on route of certification, the outcome of students placed with a particular teacher.

Reading. The reading scores of students taught by AC teachers were not significantly different from those of students of TC counterparts, as shown in the top panel of Exhibit IV.1. As the bottom panels show, the same result—no significant difference in test scores—is obtained by examining separately the comparisons of high- and low-coursework AC teachers with their TC counterparts. F-tests also confirmed that the differences between students of low-coursework AC teachers and their TC counterparts are not statistically different from the differences between the students of high-coursework AC teachers and their TC counterparts.

Although the average effect sizes (comparing student achievement in classrooms of AC teachers to that of their TC counterparts) were –0.01 and 0.00 for the low- and high-coursework subgroup analyses, effect sizes varied across mini-experiments (Exhibit IV.2). For low-coursework AC teachers, they ranged from –0.74 to 0.88 (median, –0.01). For high-coursework AC teachers, the range was –0.90 to 0.64 (median, –0.01). In 51 percent of mini-experiments, the effect size was less than zero.

[55] Kane and Staiger (2002) show that with a sample size of 25 (roughly the number of students in one mini-experiment), only about 66 percent of the variation in math scores, and 48 percent of the variation in reading scores, is due to persistent differences in quality.

[56] Hill et al. (2007) show that for K-5 students, the average gain in effect size after one year is about 0.77 in reading and 0.82 in math.

Exhibit IV.1. Spring Reading Score Differences in AC and TC Classrooms

	Number of Mini-experiments	AC Classroom Average Score	TC Classroom Average Score	Difference	Effect Size	*p*-Value
All (N=2,646)	90	38.51	38.62	–0.11	–0.01	0.84
Low Coursework (N=1,331)	48	38.29	38.50	–0.21	–0.01	0.81
High Coursework (N=1,279)	42	38.75	38.76	0.00	0.00	1.00

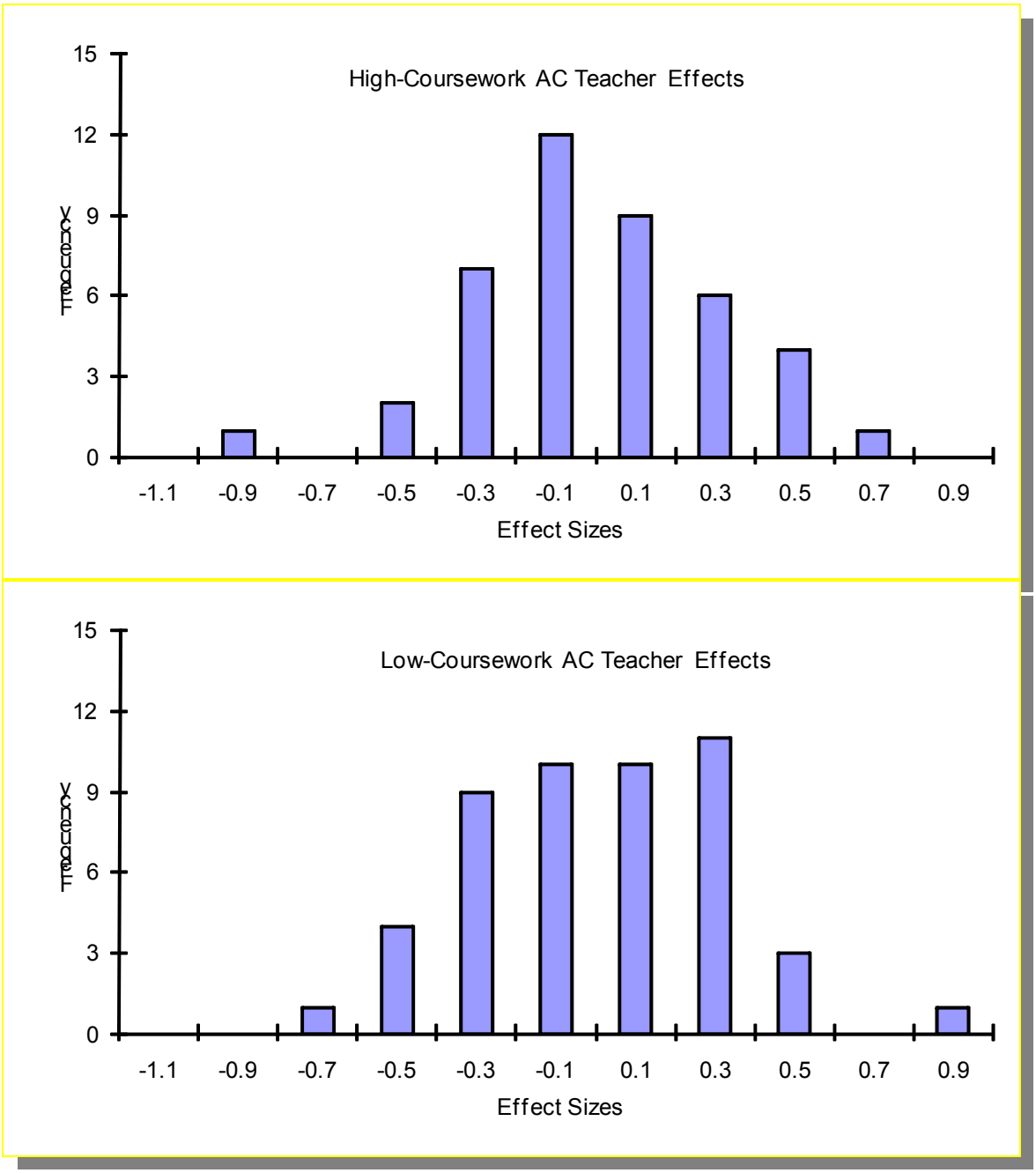
Source: California Achievement Test, 5th Edition (CAT-5), administered by Mathematica Policy Research, Inc. (MPR). Test scores are expressed in terms of normal curve equivalents (NCEs); the average score nationally is 50, and the standard deviation (SD) is 21.06.

Note: The AC classroom average score reported in the table is the TC average score plus the regression-adjusted treatment effect. The regression model controls for baseline test scores, eligibility for free or reduced-price lunch, gender, race/ethnicity, and teacher's years of experience.

Math. As shown in the top panel of Exhibit IV.3, the math scores of AC and TC students did not differ statistically. In the lower panels, the students of high-coursework AC teachers had an average NCE score of 42.0 compared to 43.5 for students of their TC counterparts, and the difference (NCEs, –1.51; effect size, –0.07) was not statistically significant at the $p<0.05$ level (Exhibit IV.3). The average difference between students of low-coursework AC teachers and students of their TC counterparts was also not statistically significant, nor was it statistically different from the difference between the students of high-coursework AC teachers and their TC counterparts.

As Exhibit IV.4 shows, effect sizes in math also varied, ranging from –1.04 to 0.64 (median, –0.03) for low-coursework teachers, and from –0.72 to 0.76 (median, –0.10) for high-coursework teachers. The effect size for students of AC teachers compared to students of their TC counterparts was less than zero in 56 percent of the mini-experiments.

Exhibit IV.2. Distribution of AC Teacher Effects in Literacy



Source: Author's calculations based on results from the CAT-5, administered by MPR.

Note: The number on the x-axis indicates the midpoint of the range of values. For example, –0.3 indicates the range of values from –0.2 to –0.4.

Exhibit IV.3. Spring Math Score Differences in AC and TC Classrooms

	Number of Mini-experiments	AC Classroom Average Score	TC Classroom Average Score	Difference	Effect Size	*p*-Value
All (N=2,578)	89	41.75	42.77	–1.01	–0.05	0.12
Low Coursework (N=1,248)	48	41.52	42.12	–0.60	–0.03	0.56
High Coursework (N=1,330)	41	42.03	43.53	–1.51	–0.07	0.10

Source: CAT-5, administered by MPR. The reading score is a total score based on vocabulary and comprehension subtests. The math test refers to the Mathematics Concepts subsection. Test scores are expressed in terms of NCEs; the average score nationally is 50, and the SD is 21.06.

Note: The AC classroom average score reported in the table is the TC average score plus the regression-adjusted treatment effect. The regression model controls for baseline test scores, eligibility for free or reduced-price lunch, gender, race/ethnicity, and teacher's years of experience.

2. Robustness Checks

The experimental analyses did not find a statistically significant relationship between selected route to certification and student test scores in reading or math, though the effects from individual mini-experiments were distributed across a range. To explore whether the overall effects reflect a peculiarity of the data or mask effects within certain groups, we checked the robustness of these findings by examining the effects on students achievement of AC teachers compared to their TC counterparts among a number of different subgroups.[57,58]

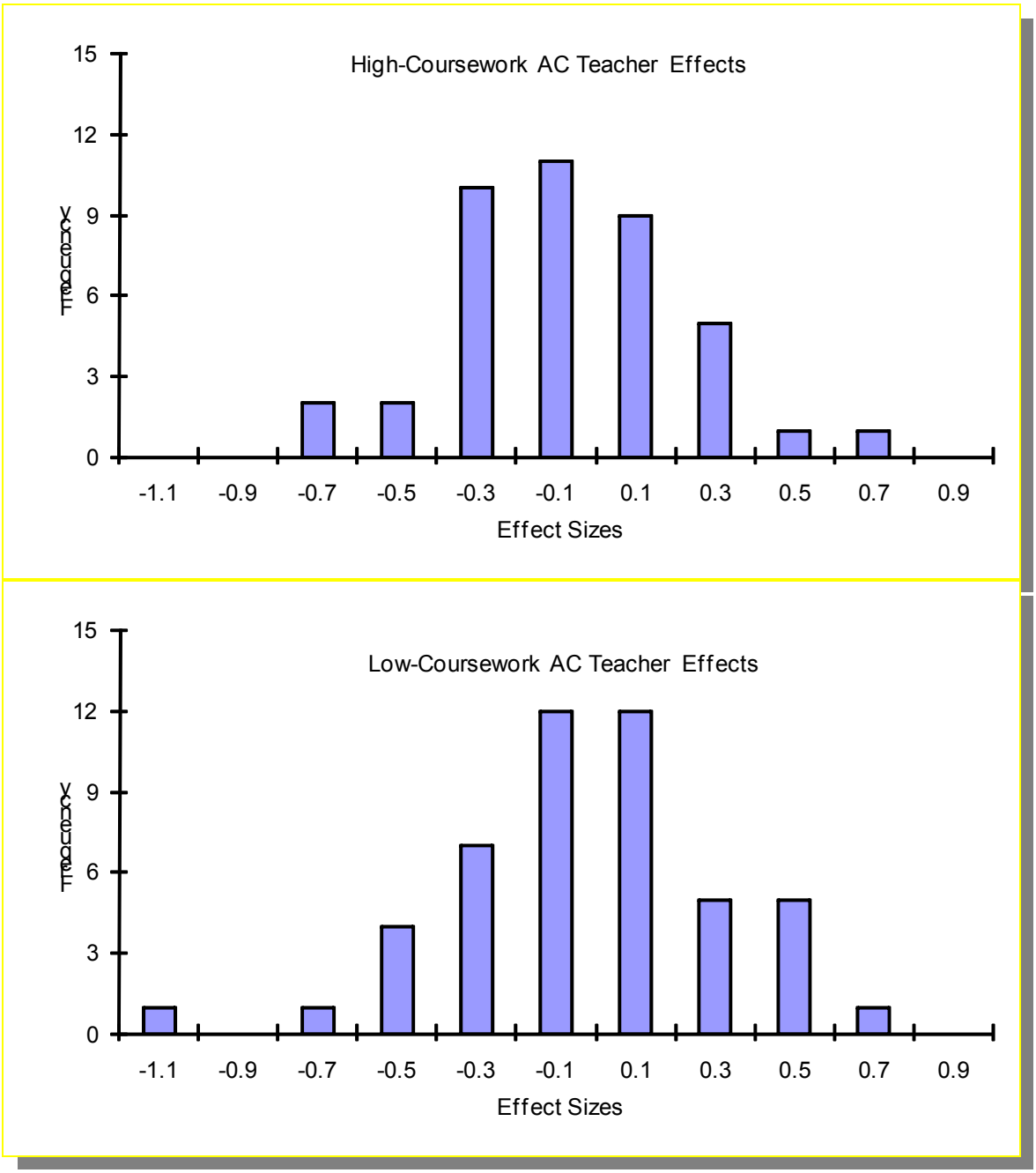
- ***State.*** Certification programs are regulated by the state, and the regulations could affect both the quality of the programs and the effectiveness of the teachers who complete them.

- ***Grade.*** The effectiveness of teachers might differ between lower and upper grades because of differences in pedagogical skills necessary to teach younger versus older elementary school students.

[57] The study was not powered to determine statistically significant differences of 0.20 SD or smaller for any subgroup containing 33 percent or less of the total sample.

[58] We also examined how sensitive the results are to alternate estimation specifications. Results from these checks are found in Appendix A, exhibits A.8 through A.11.

Exhibit IV.4. Distribution of AC Teacher Effects in Math



Source: Author's calculations based on results from the CAT-5, administered by MPR.

Note: The number on the x-axis indicates the midpoint of the range of values. For example, –0.3 indicates the range of values from –0.2 to –0.4.

- ***Teacher Experience.*** Teachers with no experience might be less effective than teachers with two or more years of experience, and this may not be entirely captured in the model that controls for experience.

- ***Coursework Status.*** AC teachers taking courses while teaching may, because of additional time demands, be less effective than those not taking courses.

State. The teachers in the study were located in seven states. We grouped every mini-experiment in the study according to the state in which the school was located. Low- and high-coursework designations fell primarily along state lines, with the exception of Texas, which had the most mini-experiments in our sample (43, with 31 low- and 12 high-coursework AC teachers), and Louisiana (6 mini-experiments with low- and 3 with high-coursework AC teachers). All AC teachers in our California sample came from high-coursework programs, and of all the states in our study, California had the most mini-experiments involving high-coursework AC teachers (21).[59] Because of the high proportion of mini-experiments in California, the effects from the high-coursework subgroups largely reflect what happened in California.[60]

The general pattern of the negative difference in math scores for students with AC teachers compared to students of their TC counterparts persists across states and is statistically significant in California, as shown in Exhibit IV.5. The relative effect on math in California is negative and statistically significant (an effect size of –0.13, nearly twice the overall effect size for high-coursework AC teachers from the basic experimental model). No other differences at the state level were statistically significant.

As illustrated in Exhibit IV.6, the negative relative effect of high-coursework AC teachers on student math achievement is restricted to California. For students of such teachers in other states, the effect (–0.01 SD) is not statistically significant.[61] Thus, the mini-experiments in California have a substantial influence on size of the overall relative effect of high-coursework AC teachers.

Grade. Coursework and other aspects of teacher training may play a greater or lesser role in the instruction received in some grades than in others. Because our sample is made up disproportionately of students in kindergarten and first grade, we divided the sample into lower grades (kindergarten and first grade) and upper grades (second through fifth).

---

[59] There were a total of 17 mini-experiments in Georgia, Illinois, New Jersey, and Wisconsin. Six included high-coursework AC teachers, and 11 included low-coursework AC teachers (of whom 10 were in New Jersey).

[60] We present the individual results for California, Louisiana, and Texas and the combined results for all other states (Georgia, Illinois, New Jersey, and Wisconsin) because of their small number of mini-experiments.

[61] An F-test of the equality of the effects across all states fails to reject the null at the $p<0.05$ level. However, because this test relied on small samples, it is not well powered.

Exhibit IV.5. Differences in Students' Spring Test Scores in AC and TC Classrooms, by State

	Number of Mini-experiments	AC Classroom Average Score	TC Classroom Average Score	Difference	Effect Size	*p*-Value
California (N = 652)[a]						
Reading	21	36.28	37.28	–0.99	–0.05	0.37
Math	20	38.85	41.76	–2.91	–0.13	0.03
Louisiana (N = 304)						
Reading	9	31.56	33.09	–1.54	–0.08	0.33
Math	9	35.89	37.90	–2.01	–0.09	0.29
Texas (N = 1,196)						
Reading	43	42.40	41.43	0.97	0.05	0.24
Math	43	45.42	46.07	–0.65	–0.03	0.52
Others (N = 458)						
Reading	17	35.10	36.09	–0.98	–0.05	0.42
Math	17	39.00	38.18	0.82	0.04	0.58

Source: CAT-5, administered by MPR. The reading score is a total score based on vocabulary and comprehension subtests. The math test refers to the Mathematics Concepts subsection. Test scores are expressed in terms of NCEs; the average score nationally is 50, and the SD is 21.06.

Note: The AC classroom average score reported in the table is the TC average score plus the regression-adjusted treatment effect. The regression model controls for baseline test scores, eligibility for free or reduced-price lunch, gender, race/ethnicity, and teacher's years of experience.

[a] One California pair was eliminated from the math analysis because the study teachers did not teach math to the students randomly assigned to their classes. This reduces the math sample sizes to 621 in California.

As Exhibit IV.7 shows, no patterns in the results suggest different effects for students in lower grades versus upper ones. Tests of the equality of coefficients indicated that no statistically significant differences existed between the lower and upper elementary grades for either high- or low-coursework AC teachers.

Teacher Experience. Disaggregating effects by years of experience reported by AC teachers showed whether differences in achievement between students of AC teachers and students of TC teachers were more pronounced among relatively inexperienced AC teachers. Previous correlational research suggests that AC teachers are less effective in their first year, but "catch up" with a year or two of experience (Boyd et al. 2006). This does not appear to be the case for the teachers in this study; the only statistically significant negative effects were found among the low-coursework AC teachers with three to four years of experience (Exhibit IV.8). Although inferences should be made with caution because of the small subgroup sizes, our sample shows no statistically significant evidence that the students of

Exhibit IV.6. Differences in Students' Spring Test Scores in AC and TC Classrooms, California and All Other States

	Number of Mini-experiments	AC Classroom Average Score	TC Classroom Average Score	Difference	Effect Size	*p*-Value
California (N = 652)[a]						
Reading	21	36.28	37.28	–0.99	–0.05	0.37
Math	20	38.85	41.76	–2.91	–0.13	0.03
All Others (N = 1,994)[a]						
Reading	69	39.19	39.02	0.16	0.01	0.81
Math	69	42.60	43.06	–0.46	–0.02	0.57
All Others—High Coursework Only (N = 626)						
Reading	21	41.23	40.22	1.00	0.05	0.34
Math	21	45.05	45.22	–0.16	–0.01	0.90

Source: CAT-5, administered by MPR. The math test refers to the Mathematics Concepts subsection. Test scores are expressed in terms of NCEs; the average score nationally is 50, and the SD is 21.06.

Note: The AC classroom average score reported in the table is the TC average score plus the regression-adjusted treatment effect. The regression model controls for baseline test scores, eligibility for free or reduced-price lunch, gender, race/ethnicity, and teacher's years of experience.

[a] One California pair was eliminated from the math analysis because the study teachers did not teach math to the students randomly assigned to their classes. This reduces the total sample sizes to 621.

novice AC teachers scored lower (relative to students of TC teachers) than students of AC teachers who had been teaching for several years.[62]

Current Coursework Status. The main coursework distinction on which this study focuses is between the AC programs with high coursework requirements and those with low ones. An alternative distinction is whether teachers are taking courses while teaching, which prior research suggests can be negatively associated with effectiveness, presumably from the multiple demands on a teacher's time (Harris and Sass 2007; Goldhaber and Anthony 2006). To test this hypothesis, we divided the sample into two groups: AC teachers who reported currently taking courses, either to complete certification requirements or to finish a degree, and those who reported not taking any courses.[63]

[62] These findings should not be interpreted as meaning that AC teachers become less effective over time. These are not longitudinal findings, but a cross-section of teachers at each level of experience. Experience levels of the TC counterparts also vary for each level of AC teacher experience.

[63] Of AC teachers in our sample, 41 percent reported that they were taking some type of coursework for certification or advanced degrees.

Exhibit IV.7. Differences in Students' Spring Test Scores in AC and TC Classrooms, by Grade Level

	Number of Mini-experiments	AC Classroom Average Score	TC Classroom Average Score	Difference	Effect Size	*p*-Value
Kindergarten–1st Grade						
Low Coursework (N = 749)						
Reading	29	32.45	33.76	–1.31	–0.07	0.24
Math	29	36.65	36.66	–0.01	0.00	0.99
High Coursework (N = 609)[a]						
Reading	21	37.36	36.55	0.81	0.04	0.52
Math	20	37.61	38.62	–1.01	–0.04	0.50
2nd–5th Grade						
Low Coursework (N = 618)[a]						
Reading	19	47.21	45.73	1.48	0.07	0.17
Math	19	48.96	50.54	–1.49	–0.07	0.26
High Coursework (N = 670)						
Reading	21	40.15	40.95	–0.80	–0.04	0.47
Math	21	46.23	48.21	–1.98	–0.09	0.15

Source: CAT-5, administered by MPR. The math test refers to the Mathematics Concepts subsection. Test scores are expressed in terms of NCEs; the average score nationally is 50, and the SD is 21.06.

Note: The AC classroom average score reported in the table is the TC average score plus the regression-adjusted treatment effect. The regression model controls for baseline test scores, eligibility for free or reduced-price lunch, gender, race/ethnicity, and teacher's years of experience.

[a] Because one pair was eliminated from the math analysis, sample sizes for math tests are smaller in this subgroup. This reduces the sample size to 578.

Students of AC teachers who reported not taking courses during the study year did not score statistically different in math or reading from students of their TC counterparts (Exhibit IV.9). In contrast, although there were no statistically significant differences in reading scores, students of AC teachers who reported taking courses scored 0.09 SD lower on their spring math tests, and the difference was statistically significant. Because 57 percent of high-coursework and 30 percent of low-coursework AC teachers in our sample reported currently taking courses, the findings based on the subgroup analyses of low- and high-coursework AC teachers were confounded with the subgroups analyses of taking courses while teaching.

Exhibit IV.8. Differences in Students' Spring Test Scores in AC and TC Classrooms, by Years of Teacher Experience

	Number of Mini-experiments	AC Classroom Average Score	TC Classroom Average Score	Difference	Effect Size	*p*-Value
1 to 2 Years of Experience						
Low Coursework (N = 668)						
Reading	24	39.65	38.11	1.55	0.09	0.29
Math	24	43.25	42.11	1.13	0.05	0.50
High Coursework (N = 463)						
Reading	15	38.77	38.46	0.32	0.02	0.84
Math	15	38.72	40.03	–1.31	–0.06	0.49
3 to 4 Years of Experience						
Low Coursework (N = 463)[a]						
Reading	17	39.09	41.66	–2.57	–0.13	0.04
Math	17	39.99	44.76	–4.77	–0.21	0.00
High Coursework (N = 483)						
Reading	15	37.46	40.18	–2.73	–0.14	0.18
Math	15	43.73	45.58	–1.86	–0.08	0.46
5+ Years of Experience						
Low Coursework (N = 75)						
Reading	3	33.64	31.24	2.40	0.12	0.51
Math	3	41.62	33.68	7.94	0.35	0.08
High Coursework (N = 333)[a]						
Reading	12	40.36	37.33	3.03	0.15	0.24
Math	11	44.23	45.51	–1.29	–0.06	0.67
Missing Experience						
Low Coursework (N = 125)						
Reading	4	30.23	32.86	–2.63	–0.13	0.45
Math	4	36.14	37.23	–1.09	–0.05	0.80

Sources: (1) CAT-5, administered by MPR. The math test refers to the Mathematics Concepts subsection. Test scores are expressed in terms of NCEs; the average score nationally is 50, and the SD is 21.06; (2) teacher survey.

Notes: No high-coursework AC teachers were missing experience. The AC classroom average score reported in the table is the TC average score plus the regression-adjusted treatment effect. The regression model controls for baseline test scores, eligibility for free or reduced-price lunch, gender, race/ethnicity, and teacher's years of experience as a classroom teacher of record.

[a] Because one pair was eliminated from the math analysis, sample sizes for math tests are smaller in this subgroup. This reduces the sample size to 302.

Exhibit IV.9. Differences in Students' Spring Test Scores in AC and TC Classrooms, by Whether the AC Teacher Is Currently Taking Courses

	Number of Mini-experiments	AC Classroom Average Score	TC Classroom Average Score	Difference	Effect Size	*p*-Value
Taking Courses (N = 877)						
Reading	37	37.49	38.03	–0.54	–0.03	0.50
Math	37	39.95	42.03	–2.08	–0.09	0.04
Not Taking Courses (N = 1,769)						
Reading	53	39.92	39.03	0.20	0.01	0.77
Math	52	43.04	43.29	–0.26	–0.01	0.76

Sources: (1) CAT-5 and teacher survey responses, both administered by MPR. The math test refers to the Mathematics Concepts subsection. Test scores are expressed in terms of NCEs; the average score nationally is 50, and the SD is 21.06; (2) teacher survey.

Note: The AC classroom average score reported in the table is the TC average score plus the regression-adjusted treatment effect. The regression model controls for baseline test scores, eligibility for free or reduced-price lunch, gender, race/ethnicity, and teacher's years of experience. Five low-coursework AC teachers did not answer the survey. These were treated as not taking coursework.

[a] Because one pair was eliminated from the math analysis, sample sizes for math tests are smaller in this subgroup. This reduces the sample size to 1,738 for the "not taking courses" group.

Sixty-two percent of California AC teachers reported taking courses, versus 35 percent of AC teachers outside California. We examined whether the students of the California AC teachers who were taking courses scored lower than the students of the ones who were not, and found that students of those taking coursework scored lower than students of their TC counterparts (effect size, –0.16; p=0.03), while students of California AC teachers not taking courses had scores not statistically significantly different from those of students of the TC counterparts.

3. Teacher Practices

Though there were no statistically significant differences in student test scores in reading or math, there may be differences in the classroom practices experienced by students of teachers trained through different routes. The instruction that students receive in the classroom may influence how students learn, which may not be fully captured through test scores immediately following the intervention year. One way we measured the instruction received by students in the classrooms was to conduct observations using the Vermont Classroom Observation Tool (VCOT), which, as discussed in Chapter II, assesses teachers in three domains: (1) implementation of the lesson, (2) content of the lesson, and (3) culture of the classroom in which the lesson was conducted. We used a 5-point scale to rate indicators in each of these areas, observing each teacher in the study during two math and two literacy instruction periods over the course of four days. The domain scores were averaged for the two observations in each subject area. The summary statistics for the VCOT for all study

teachers are in Exhibit IV.10. On average, VCOT scores for teachers in the analysis sample were lowest in math content (1.54) and highest in literacy culture (2.92). SD range was 0.7 to 0.87.[64]

Exhibit IV.10. Descriptive Statistics of Vermont Classroom Observation Tool Scores

	Average	Standard Deviation
VCOT Literacy (N = 184)		
Content	2.28	0.70
Culture	2.92	0.78
Implementation	2.52	0.77
VCOT Mathematics (N = 182)		
Content	1.54	0.76
Culture	2.86	0.87
Implementation	2.42	0.79

Source: Ratings are from the Vermont Classroom Observation Tool (VCOT) and range from 1 to 5 for each domain.

The instruction experienced by students of AC teachers overall was not statistically different from the instruction experienced by students of their TC counterparts in literacy (Exhibit IV.11) or math (Exhibit IV.12), as measured by the VCOT. In the subgroup analyses, the classroom instruction was similar for the AC teachers from low-coursework programs and their TC counterparts. The classes of AC teachers from high-coursework programs, however, were statistically different from those of their TC counterparts on one of the measures: classroom culture in literacy. The remaining VCOT measures of classroom practices were not statistically different between the classrooms of high-coursework AC teachers and their TC counterparts.[65,66] The average scores within each dimension were consistently similar in magnitude and did not differ statistically among the classrooms of high-coursework AC teachers, low-coursework AC teachers, and the TC counterparts to the low-coursework teachers. In contrast, the TC teachers matched to high-coursework AC teachers were rated statistically higher than all other teachers in literacy culture (p=0.01), literacy implementation (p=0.03), and math content (p=0.02). Thus, although TC teachers matched with low- and high-coursework AC teachers were required to

[64] Contractors were trained to administer the VCOT observation. During training, they observed and scored a videotaped class, and their 16-item scores were compared to the scores of an expert panel consisting of the tool's developer and two trained observers who demonstrated high rates of agreement in scoring. Trainees had two opportunities to come within 0.75 points of the panel's average score for each of the three constructs (implementation, content, and culture) during a test observation. Trainees who did not meet the standard were not allowed to conduct observations. See Appendix A for more details.

[65] To calculate effect sizes, we took the estimated effect and divided by the SD of the TC teachers' classroom measure. The effect size calculation used the same SD for high- and low-coursework AC teachers.

[66] The sample size for VCOT observations allowed us a minimum detectable effect size of 0.37 with 80 percent power.

complete a similar amount of coursework, this finding suggests that TC teachers matched with high-coursework AC teachers differed in some other way.[67]

Exhibit IV.11. Differences Classroom Practices in Literacy

	Number of Mini-experiments	AC Teacher Average Score	TC Teacher Average Score	Difference	Effect Size	*p*-Value
All (N = 184)						
Content	87	2.25	2.29	–0.04	–0.06	0.69
Culture	87	2.81	2.97	–0.16	–0.21	0.15
Implementation	87	2.44	2.55	–0.11	–0.14	0.32
Low Coursework (N = 99)						
Content	46	2.27	2.20	0.07	0.10	0.64
Culture	46	2.75	2.77	–0.02	–0.03	0.89
Implementation	46	2.44	2.39	0.05	0.07	0.74
High Coursework (N = 85)						
Content	39	2.27	2.41	–0.15	–0.22	0.28
Culture	39	2.75	3.21	–0.31	–0.40	0.03
Implementation	39	2.44	2.76	–0.29	–0.37	0.07

Source: Ratings are from the VCOT and range from 1 to 5 for each domain.

Note: The AC average score reported in the table is the TC average score plus the regression-adjusted AC mean. The regression model controls for the teacher's years of experience.

4. Summary of Experimental Findings

The study includes comparisons of teachers who chose AC routes to certification to teachers who chose TC routes, and the findings from the experimental analyses indicate that there was no statistically significant difference in student achievement from placing an AC teacher in the classroom when the alternative was a TC teacher. The average differences in effects were also not statistically significant across the two AC subgroups (low- or high-coursework). This is evidence that a student's performance on standardized tests, *on average,* is expected to be the same regardless of whether a classroom is headed by a TC or an AC teacher. However, effects varied across all teachers, with effect sizes ranging from –1.0 to 0.9; this translates to differences in achievement of more than one grade level. The variation provides an estimate of uncertainty about whether placing an AC teacher rather than a TC teacher in the classroom will lead to *differences in* student standardized achievement scores.

[67] This is a descriptive finding based on the VCOT observations data, not a finding based on the experimental design of the study.

Exhibit IV.12. Differences Classroom Practices in Mathematics

	Number of Mini-experiments	AC Teacher Average Score	TC Teacher Average Score	Difference	Effect Size	*p*-Value
All (N = 180)						
Content	86	1.44	1.55	–0.11	–0.15	0.31
Culture	86	2.77	2.88	–0.11	–0.13	0.37
Implementation	86	2.34	2.43	–0.09	–0.11	0.44
Low Coursework (N = 83)						
Content	46	1.39	1.38	0.01	0.01	0.95
Culture	46	2.71	2.68	0.03	0.03	0.88
Implementation	46	2.39	2.29	0.10	0.12	0.55
High Coursework (N = 97)						
Content	40	1.52	1.78	–0.25	–0.33	0.10
Culture	40	2.90	3.14	–0.24	–0.27	0.17
Implementation	40	2.35	2.61	–0.26	–0.33	0.12

Source: Ratings are from the VCOT and range from 1 to 5 for each domain.

Note: The AC average score reported in the table is the TC average score plus the regression-adjusted AC mean. The regression model controls for the teacher's years of experience.

Teacher performance based on a measure of practice provides another way to compare teachers hired from the different routes. Of six measured ratings of classroom practices between AC teachers and their TC counterparts, one showed a statistically significant difference. This further suggests that information about the route a teacher chooses to acquire certification does not predict performance in the classroom.

B. NONEXPERIMENTAL ANALYSES

The average classroom effects estimated in the experimental analysis were not statistically different from zero. However, effects varied across pairs of teachers, as shown in exhibits IV.2 and IV.4, and students of AC teachers scored higher than students of their TC counterparts in nearly as many cases as they scored lower (49 percent in reading and 44 percent in math).[68] As indicated in chapters II and III, the extent and nature of preparation by teachers in the sample vary by route; the characteristics and experiences of those who select AC and TC also vary. Thus, the relative differences in student outcomes could be explained by their background characteristics, skills they gained during their training, other factors, or some combination. Because teachers were not randomly assigned to their training programs, experimental methods cannot separate the effects of teacher characteristics from the influence of their training. Similarly, the experimental methods cannot determine whether classroom practices contribute to student achievement.

[68] We tested the distribution of the effects using the "Q-test" suggested by Lipsey and Wilson (2001), which tests whether the observed variance in the estimates is greater than would be expected from sampling error alone. We reject the hypothesis of homogeneous effects.

In this section, we use nonexperimental methods to examine whether differences in observable teacher characteristics, training, transitional support experiences, and classroom practices are associated with the teacher-level effects. Each in turn has the potential to inform the improvement of teacher quality. Correlational evidence about the relationship between training experiences and classroom practices, as well as support experiences and effectiveness in the classroom, provide suggestive information relevant to the structure and content of teacher preparation programs. We focus on differences between AC teachers and their TC counterparts to measure whether the differences in background and training characteristics between the two teacher types explained the variation in estimated effects across the mini-experiments. These analyses are nonexperimental because teachers were not randomly assigned to teacher programs. Thus, the findings are suggestive and cannot be interpreted as causal.

To estimate the nonexperimental correlations, we used the following model

$$(1)\quad y_{ijk} = \beta_0 + \beta_x X_{ijk} + \lambda_0 AC_{jk} + \lambda_1 dZ_k AC_{jk} + AC_{jk}\nu_k + \varepsilon_{ijk}$$

where dZ_k is the difference between the AC and TC teacher in school k in some characteristic (such as hours of instruction in a particular subject area or SAT score) and ν_k is an unobserved random variable. In words, this equation estimates the correlation between a student's posttest score and student-level characteristics (including pretest score), whether his or her teacher was from an AC program, differences between the characteristics of AC and TC teacher pair within a school and grade, and other unobservable effects. The coefficient of interest is λ_1, which provides a measure of the correlation of the differences between AC and TC teachers and the students' posttest scores.

We estimated equation (1) using ordinary least squares, with clustering accounted for in the standard errors using the Huber-White sandwich estimator. Because our goal was to explain effects with observable differences between the teachers, we restricted the sample to teachers who did not leave the sample during the study. We applied this restriction because it is unclear what relationship the teacher's characteristics had in the overall outcome when the teacher taught for only a portion of the year.

1. Differences in the Amount of Coursework

The findings from the experimental analysis showed no statistically significant evidence that students of teachers from high-coursework AC programs scored higher relative to the students of TC teachers than students of teachers from low-coursework AC programs, which suggests that the amount of coursework required in AC programs does not make a difference in student achievement.[69] However, the experimental estimates are not ideal for isolating the effects of required coursework. Although similar to each other on average, the

[69] Those results included students whose [0]teachers who left the sample during the study, so differential teacher attrition may have influenced the results. However, the same analysis excluding students and teachers who left their assigned classrooms similarly failed to find any evidence of a positive relationship between the amount of teacher coursework and student outcomes (Appendix A, Exhibit A.10).

TC comparison teachers also vary in the amount of coursework required in their training programs and do not provide a completely consistent benchmark to measure against. Second, the nonexperimental analysis is also not sufficient to measure a causal effect, because the teachers are not randomly assigned to programs with different levels of coursework. However, the nonexperimental framework is useful for determining whether there is a correlation between teacher coursework and student outcomes that merits further investigation.

For the nonexperimental analysis, we categorized the hours of required coursework into the topics that might be expected to influence performance in teaching reading and math: reading/
language arts pedagogy, math pedagogy, classroom management, child development, student assessment, and a residual "other" category for miscellaneous education-related coursework. Hours of fieldwork were also accounted for. Then we estimated regressions relating hours of required coursework to student test scores and found that the number of required hours did not have a statistically significant correlation with student outcomes in either reading or math.[70] Results from these regressions are shown in Appendix A, Exhibit A.12.

2. Differences in Education and Support Experiences

Just as required coursework varies between AC and TC programs, other experiences may also vary. As shown in Chapter III, the AC teachers in our sample were less likely than their TC counterparts to report having majored in education, more likely to report having been engaged in coursework while teaching, and more likely to report having a mentor during their first year. These differences may lead to differences in student outcomes. We examined the relationship between student achievement and the following education and support experiences: (1) master's degree, (2) undergraduate college major, (3) formal mentoring in first year of teaching, (4) regular opportunities to observe other teachers in first year, (5) regular supportive communication with school officials in first year of teaching, and (6) currently taking courses toward certification or a higher degree.

Two measures were related to student outcomes at a statistically significant level, negatively in each case. The students of AC teachers with master's degrees had lower reading scores (effect size –0.13) than the students of the TC counterparts without a master's degree. These findings are consistent with prior literature that typically fails to find any positive relationship between master's degrees and student achievement (Hanushek 1997).[71] Similarly, students of AC teachers who reported taking courses scored lower in reading

[70] Some of the characteristics included in the regressions may be highly correlated with each other. To allow for this, we entered each variable one at a time into a regression to determine whether any were statistically significant in isolation. This approach generated findings similar to those obtained when using a group of variables jointly to estimate the regressions.

[71] In a comprehensive review of the education literature that seeks to measure the impacts of individual teachers on student achievement, Hanushek (1997) was unable to find any studies that found a positive and significant relationship between a teacher's education level and student outcomes; however, he did identify 10 studies that found a negative and statistically significant relationship.

(effect size –0.12) than did students of their TC counterparts who reported not taking coursework. This finding reinforces the subgroup findings presented in the previous section.[72] The results are shown in Appendix A, Exhibit A.13.

3. Differences in Teacher Characteristics

Teachers in our sample who attended AC programs differed in a number of personal characteristics from those who attended TC programs. Prior research has shown that some teacher characteristics, such as cognitive ability, have an important effect on student achievement (Goldhaber 2006; Ferguson and Ladd 1996; Ferguson 1991). Therefore, differences in characteristics between AC and TC teachers may explain some of the variation in student outcomes. For these estimates, we included all measured teacher characteristics that might be correlated with student outcomes: SAT score, whether the teacher attended a selective undergraduate institution, race/ethnicity, gender, and age. None of the differences between AC and TC teachers on any of these dimensions correlated with the outcomes of their students. The results are shown in Appendix A, Exhibit A.14.

The AC teachers in the study were statistically significantly more likely than their TC counterparts to be black (35 percent versus 11 percent), but not to be Hispanic. Since many of the students in the study were also black or Hispanic (35 percent and 47 percent, respectively), students with an AC teacher were more likely to be matched with a teacher of their own race/ethnicity. Nearly half (49 percent) of black students in AC classrooms had a black teacher, while 18.4 percent of black students in a TC classrooms had a black teacher. The difference for Hispanic students was not as large (38.4% of Hispanic students in an AC classroom had a Hispanic teacher, compared to 26.7% in a TC classroom), but the difference was also statistically significant. While there is no evidence that the race/ethnicity of a teacher is related to student achievement in general, experimental and nonexperimental research has shown either no effect or a positive and statistically significant effect on student achievement when African American students are matched with a teacher of the same race (Ehrenberg and Brewer 1995; Ehrenberg et al. 1995; Dee 2004; Clotfelter et al. 2007). If having a teacher of the same race/ethnicity has a positive impact on achievement, overall differences in teacher race/ethnicity may not capture the benefits that accrue to a particular subgroup of the student population. The random assignment of students for this study allowed us to examine whether the black and Hispanic students who were matched to a teacher of the same race/ethnicity performed better than black and Hispanic students who were not matched.[73] Exhibit A.15 in Appendix A displays the results from these regressions.

[72] One key difference between the subgroup findings and these is that the subgroup findings do not account for whether the TC teacher is taking coursework, though some TC teachers are working toward an advanced degree. When the subgroup analysis from Exhibit IV.9 is restricted to those instances in which AC teachers are taking courses and TC teachers are not, the effect size is –0.06 (p=0.23) for reading and –0.11 (p=0.04) for math. This suggests that TC teachers who are taking coursework may have effects different from those of AC teachers who are taking coursework.

[73] The specification for this model differed from what is shown in equation 1 to allow more direct interpretation of the coefficient on student-teacher racial/ethnic match. This model is specified as

The coefficients, which represent the NCE difference in test scores for students matched with a teacher of the same race/ethnicity, do not statistically differ (NCE differences for black students are 3.45 in math, p=0.09 and 2.35 in reading, p=0.17; NCE differences for Hispanic students are –1.43 in math, p=0.57 and –2.52 in reading, p=0.40).

4. Differences in Teacher Practices

Differences in teacher practices may be associated with overall differences in student achievement. A fourth set of regressions examined the relationship between the teachers' VCOT observation ratings and the relative effects on student achievement. In terms of statistical significance, none of the differences were positively related to relative effects, and the score for classroom culture in literacy was negatively related to reading scores (shown in Appendix A, Exhibit A.16). This implies that AC teachers who scored higher than their TC counterparts on classroom culture when teaching literacy had students with lower reading scores. Overall, the lack of a statistically significant relationship between observation ratings and student achievement suggests that differences in practices between high-coursework AC teachers and their TC counterparts were not associated with student achievement.

As described in Chapter II, principals rated teachers compared to other teachers in the school, with a value of 3.0 indicating that the teacher was average compared with the others. However, principals were not blind to the teacher's research status the way the VCOT observers were. Similarly, there is no guarantee that the principals were rating the teachers' performance in the experimentally constructed classrooms (that is, principals could be providing an overall impression of the teachers gained through multiple years of interactions). For these reasons, the principal ratings cannot be considered experimental outcomes, as the VCOT scores are. However, they provide insight into how principals perceive teachers as opposed to how independent observers rate them. Principals may also be able to detect teacher attributes or practices that influence student achievement.

The ratings that principals provided were grouped into three categories: reading/language arts instruction, math instruction, and classroom management. Average principal scores ranged from 3.7 to 3.9 (Exhibit IV.13); SD range was 0.82 to 0.96. The

---

(continued)

$$y_{ijk} = \beta_0 + \lambda_0 AC_{jk} + \lambda_1 BTea * BStu_{ijk} + \lambda_2 HTea * HStu_{ijk} + \lambda_3 BStu_{ijk} + \lambda_4 HStu_{ijk} + \lambda_5 BTea_{jk} + \lambda_6 HTea_{jk} + \beta_x X_{ijk} + \nu_j + \varepsilon_{ijk}$$

Where *BTea*BStu* is dummy variable indicating that a student is a black student in the class of a black teacher, *HTea*HStu* is a dummy indicating that the student is a Hispanic student in the class of a Hispanic teacher, *BTea* is an indicator that the teacher is black, *HTea* is an indicator that the teacher is Hispanic, *BStu* is an indicator that the student is black, *HStu* is an indicator that the student is Hispanic, *X* is a vector of other student characteristics (baseline test scores in all subjects, race, gender, and free/reduced-price lunch status), and *i,j,k* index student, class, and pair. The coefficient λ_1 is the marginal effect for a black student assigned to the class of a black teacher, and λ_2 is the marginal effect for a Hispanic student assigned to the class of a Hispanic teacher.

range of scores indicates that, on the whole, principals rated the study teachers above average compared to other teachers in their schools.

There were no statistically significant differences in principals' ratings between AC and TC teachers and their TC counterparts, nor were there differences for the low- or high-coursework subgroups. In contrast to the VCOT findings, there were no statistically significant differences between principals' ratings of the TC teachers to whom the high coursework AC teachers were matched and the rest of the sample. However, this comparison should be interpreted with caution, since the principals, unlike the VCOT observers, were not trained to be or expected to be consistent across settings.

Exhibit IV.13. Descriptive Statistics of Principals' Ratings of Teachers' Performance

	Average	Standard Deviation
Reading/Language Arts	3.73	0.83
Math	3.72	0.83
Classroom Management	3.92	0.87

Exhibit IV.14. Differences in Principal Ratings of Classroom Practices

	Number of Mini-experiments	AC Teacher Average Rating	TC Teacher Average Rating	Difference	Effect Size	*p*-Value
All AC (N = 188)						
Reading/Language Arts	90	3.63	3.84	–0.20	–0.26	0.09
Math	90	3.66	3.78	–0.15	–0.15	0.32
Classroom Management	91	3.84	4.01	–0.23	–0.23	0.17
Low Coursework (N = 101)						
Reading/Language Arts	48	3.59	3.82	–0.23	–0.30	0.16
Math	48	3.57	3.70	–0.13	–0.15	0.42
Classroom management	48	3.88	3.96	–0.08	–0.11	0.59
High Coursework (N = 87)						
Reading/Language Arts	42	3.63	3.86	–0.23	–0.30	0.21
Math	42	3.69	3.87	–0.17	–0.21	0.35
Classroom management	42	3.75	4.07	–0.32	–0.42	0.08

Source: Principal interviews. Ratings range from 1 to 5 for each domain.

We also examined the correlational relationship between principal ratings of teacher practices and student achievement using the model presented in equation 1 above. That is, we included the difference in the principal rating between an AC teacher and the TC counterpart as an explanatory variable in the model (results shown in Appendix A, Exhibit A.17). The one statistically significant relationship was that AC teachers who were rated higher in classroom management than their TC counterparts have students who scored *lower* on reading exams than their counterparts. Classroom management included some indicators similar to classroom culture using the VCOT, so this finding is consistent with the VCOT finding described earlier. As with the VCOT observation ratings, this counterintuitive finding suggests that principals were assigning above-average ratings to practices that were not positively correlated with student achievement.

5. Summary of Nonexperimental Findings

As previously indicated, AC and TC teachers in the study sample differed along various dimensions. Some of the differences arose from the type of training or support they received; others were indications of the types of people who enrolled in these two kinds of programs. In total, differences in AC teachers' characteristics explained about 5 percent of the variation in effects on math test scores and less than 1 percent on reading test scores.[74] We found no evidence that differences in the amount or content of coursework were related to a teacher's performance. However, students of AC teachers currently taking coursework scored statistically lower on reading tests. Differences in other background characteristics, such as demographics or educational attainment, and in other measures of cognitive ability, did not explain the variation on effects across teachers.

In general, these findings are similar to those from other research that indicates that the variation in the effectiveness of teachers in improving student test scores is not easily explained by observable training or teacher characteristics (Rivkin et al. 2005).

C. SUMMARY

The analyses in this report were designed to examine the relative effectiveness of AC and TC teachers and to explore teacher characteristics and aspects of teacher training that are associated with student achievement. Key findings from the empirical analyses include:

- ***There was no statistically significant difference in performance between students of AC teachers and those of TC teachers.*** Average differences in reading and math achievement in all instances were not statistically significant. Furthermore, students of AC teachers scored higher than students of their TC counterparts in nearly as many cases as they scored lower (49 percent in reading and 44 percent in math). The effects of AC teachers varied across experiments, and nonexperimental correlations explained 5 percent of the variation in math

[74] These percentages are calculated from a single model that includes differences in all observable characteristics. Rivkin et al. (2005) estimate that at least 7.5 percent of variation in student achievement is due to teacher quality, though very little of this can be explained with observable teacher characteristics like those included in this analysis.

and 2 percent in reading. Therefore, the route to certification selected by a prospective teacher is unlikely to provide information, on average, about the expected quality of that teacher in terms of either classroom practices or student achievement.

- ***There is no evidence from this study that greater levels of teacher training coursework were associated with the effectiveness of AC teachers in the classroom.*** Treating the students of TC teachers as a common benchmark, the experimental results provided no evidence that the students of low-coursework AC teachers scored statistically different from those of their TC counterparts, nor did students of high-coursework AC teachers compared to their TC counterparts. Correlational analysis similarly failed to show that the amount of coursework was associated with student achievement. Therefore, there is no evidence that AC programs with greater coursework requirements produce more effective teachers.

- ***There is no evidence that the content of coursework is correlated with teacher effectiveness.*** After controlling for other observable characteristics that may be correlated with a teacher's effectiveness, there was no statistically significant relationship between student test scores and the content of the teacher's training, including the number of required hours of math pedagogy, reading/language arts pedagogy, or fieldwork. Similarly, there was no evidence of a statistically positive relationship between majoring in education and student achievement.

- ***Students of AC teachers who were taking coursework while teaching scored lower in math than students of their TC counterparts.*** The students of AC teachers taking coursework scored an average of 40.17 on their math tests, compared with an average score of 42.25 for the students of their TC counterparts (p=0.04). This finding suggests that student performance in an AC teacher's class may be negatively related to the teacher's taking courses while teaching.

REFERENCES

Ballou, D., and M. Podgursky. Teacher Pay and Teacher Quality. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research, 1997.

Boyd, D., P. Grossman, H. Lankford, S. Loeb, and J. Wyckoff. "How Changes in Entry Requirements Alter the Teacher Workforce and Affect Student Achievement." National Bureau of Economic Research (Working Paper No. 11844), 2005.

California Achievement Tests, Fifth Edition: Technical Report. Monterey, CA: CTB/ McGraw-Hill, 2006.

Corcoran, S.P., W.N. Evans, and R.S. Schwab. "Women, the Labor Market, and the Declining Relative Quality of Teachers." Journal of Policy Analysis and Management, vol. 23, no. 3, 2004, pp. 449–470.

Danielson, C. Enhancing Professional Practice: A Framework for Teaching. Alexandria, VA: Association for Supervision and Curriculum Development, 1996.

Darling-Hammond, L. "Teaching and Knowledge: Policy Issues Posed by Alternative Certification for Teachers." Peabody Journal of Education, vol. 67, no. 3, 1992, pp. 123–154.

Decker, P.T., D.P. Mayer, and S. Glazerman. "The Effects of Teach For America on Students: Findings from a National Evaluation." Princeton, NJ: Mathematica Policy Research, Inc., June 2004.

Decker, P.T., J.G. Deke, A.W. Johnson, D.P. Mayer, J. Mullens, and P.Z. Schochet. "The Evaluation of Teacher Preparation Models: Design Report." Princeton, NJ: Mathematica Policy Research, Inc., October 2005.

Dee, T.S. "Teachers, Race and Student Achievement in a Randomized Experiment." Review of Economics and Statistics, vol. 86, no. 1, 2004, pp. 195–210.

Dwyer, C.A. Development of the Knowledge Base for the Praxis III: Classroom Performance Assessments Assessment Criteria. Princeton, NJ: Educational Testing Service, 1994.

Feistritzer, C.E., and D.T. Chester. "Alternative Teacher Certification: A State-by-State Analysis." Washington, DC: National Center for Education Information, 2002.

Ferguson, R.F. "Paying for Public Education: New Evidence on How and Why Money Matters." Harvard Journal on Legislation, vol. 28, 1991, pp. 465–498.

Ferguson, R.F., and H.F. Ladd. "How and Why Money Matters: An Analysis of Alabama Schools." In Holding Schools Accountable: Performance-Based Reform in Education, edited by H. Ladd. Washington, DC: Brookings Institution, 1996, pp. 265–298.

Finn, C.E. "High Hurdles." Education Next, vol. 3, no. 2, 2002, pp. 62–67.

Goldhaber, D.D., and E. Anthony. "Can Teacher Quality Be Effectively Assessed? National Board Certification as a Signal of Effective Teaching." Review of Economics and Statistics, vol. 89, no. 1, 2007, pp. 134–150.

Goldhaber, D.D., and D.J. Brewer. "Why Don't Schools and Teachers Seem to Matter? Assessing the Impact of Unobservables on Educational Productivity." Journal of Human Resources, vol. 32, no. 3, 1997, pp. 505–523.

Goldhaber, D.D., and D.J. Brewer. "Does Teacher Certification Matter? High School Teacher Certification Status and Student Achievement." Educational Evaluation and Policy Analysis, vol. 22, no. 2, 2000, pp. 129–145.

Hanushek, E.A. "Assessing the Effects of School Resources on Student Performance: An Update." Educational Evaluation and Policy Analysis, vol. 19, no. 2, 1997, pp. 141–164.

Hanushek, E.A., J.F. Kain, D.M. O'Brien, and S.G. Rivkin. "The Market for Teacher Quality." National Bureau of Economic Research (Working Paper no. 11154), 2005.

Hanushek, E.A., S.G. Rivkin, and J.F. Kain. "Why Public Schools Lose Teachers." Journal of Human Resources, vol. 39, no. 2, 2004, pp. 326–354.

Harris, D.N., and T.R. Sass. "Teacher Training, Teacher Quality and Student Achievement." Working paper, February 2007.

Hess, F.M. "Tear Down This Wall: The Case for a Radical Overhaul of Teacher Certification." Washington, DC: Progressive Policy Institute, November 2001.

Jelmberg, J. "College-Based Teacher Education Versus State-Sponsored Alternative Programs." Journal of Teacher Education, vol. 47, no. 1, 1996, pp. 60–66.

Kane, T.J., J.E. Rockoff, and D.O. Staiger. "What Does Certification Tell Us About Teacher Effectiveness? Evidence from New York City." National Bureau of Economic Research (Working Paper No. 12155), 2006.

Laczko-Kerr, I., and D.C. Berliner. "The Effectiveness of 'Teach For America' and Other Under-Certified Teachers on Student Academic Achievement: A Case of Harmful Public Policy." Education Policy Analysis Archives, vol. 10, no. 37, 2002.

Lutz, F.W., and J.B. Hutton. "Alternative Teacher Certification: Its Policy Implications for Classroom and Personnel Practice." Educational Evaluation and Policy Analysis, vol. 11, no. 3, 1989, pp. 237–254.

Mayer, D.P., P.T. Decker, S. Glazerman, and T.W. Silva. "Identifying Alternative Certification Programs for an Impact Evaluation of Teacher Preparation." Cambridge, MA: Mathematica Policy Research, Inc., April 2003.

Miller, J.W., M.C. McKenna, and B.A. McKenna. "A Comparison of Alternatively and Traditionally Prepared Teachers." Journal of Teacher Education, vol. 49, no. 3, 1998, pp. 165–176.

Monk, D. "Subject Area Preparation of Secondary Math and Science Teachers and Student Achievement." Economics of Education Review, vol. 13, no. 2, 1994, pp. 125–145.

Monk, D.H., and J.K. King. "Multilevel Teacher Resource Effects on Pupil Performance in Secondary Mathematics and Science: The Case of Teacher Subject-Matter Preparation." In Choices and Consequences: Contemporary Policy Issues in Education, edited by R. Ehrenberg. Ithaca, NY: ILR Press, 1994, pp. 29–58.

Peterson, P.E., and D. Nadler. "What Happens When States Have Genuine Alternative Certification?" Education Next, vol. 9, no 1, 2009, pp. 70–74.

National Institute of Child Health and Human Development. "Report of the National Reading Panel: Teaching Children to Read Reports of the Subgroups." 2000. Available online at [http://www.nichd.nih.gov/publications/nrp/report.htm]. Accessed July 21, 2006.

Raymond, M., S.H. Fletcher, and J. Luque. Teach For America: An Evaluation of Teacher Differences and Student Outcomes in Houston, Texas. Stanford, CA: CREDO, Stanford University, 2001.

Rivkin, S.G., E.A. Hanushek, and J.F. Kain. "Teachers, Schools and Academic Achievement." Econometrica, vol. 73, no. 2, 2005, pp. 417–458.

Stoddard, C. "Why Has the Number of Teachers per Student Risen While Teacher Quality Has Declined? The Role of Changes in the Labor Market for Women." Journal of Urban Economics, vol. 53, no. 3, 2003, pp. 458–481.

U.S. Department of Education. Elementary and Secondary Education: Key Policy Letters Signed by the Education Secretary or Deputy Secretary. Washington, DC: U.S. Department of Education, October 21, 2005. Available online at [http://www.ed.gov/policy/elsec/guid/secletter/051021.html]. Accessed September 24, 2007.

Walsh, K., and S. Jacobs. Alternative Certification Isn't Alternative. Washington, DC: Thomas B. Fordham Institute, 2007.

Zeichner, K.M., and A. Schulte. "What We Know and Don't Know from Peer-Reviewed Research About Alternative Teacher Certification Programs." Journal of Teacher Education, vol. 52, no. 4, 2001, pp. 266–282.

APPENDIX A

SUPPLEMENTARY TECHNICAL INFORMATION ON DATA COLLECTION, RESPONSE RATES, AND ANALYSIS

A. IMPLEMENTING RANDOM ASSIGNMENT

Random assignment was conducted during spring and summer of 2004 and 2005, prior to the start of the next school year.[75] Schools provided student rosters, and we randomly assigned students to the study teachers. We accommodated the following requests by schools while maintaining the integrity of random assignment:

- We allowed principals to exclude a small number of students (not more than 10 percent of any class) from random assignment,[76] such as a student who was being retained in a grade and had to be placed with a specific study teacher.

- We honored principals' requests to separate some students from one another. For example, students might be separated because they were siblings or because they did not get along well together. In these cases (about 3 percent of the total sample), we would randomly assign the students to different teachers.

- We used stratified random assignment in 95 percent of sites to balance classrooms with respect to characteristics of concern to school staff. In most cases (90 percent), we stratified by gender and one or two additional student characteristics, such as academic ability or race/ethnicity. This accommodation was also useful for the study, since it reduced the chance of random imbalance in the makeup of classes.

After school began, we confirmed that random assignment had been maintained. One to two weeks into the fall semester, we verified with school officials that the rosters we created for the study classrooms were actually being used for those classes. Very few students (less than 3 percent) were originally assigned to one type of study teacher but switched over to the other type (crossovers). Because this percentage was small, we did not correct for it in the analysis. The results remain unchanged if the crossover students are eliminated.

B. DATA COLLECTION ON STUDENTS AND TEACHERS

1. Student Achievement Tests

We obtained information on students' reading and math achievement by administering the commercially available and widely used California Achievement Test, 5th Edition (CAT-5). We conducted baseline testing in reading and math a few weeks after the start of each school year, and follow-up testing a few weeks before the end of the school year. Kindergartners took the Complete Battery, Level K; first graders took the Complete Battery, Level 10; second graders, the Complete Battery, Level 11; third graders, the Survey Battery,

[75] Random assignment occurred as late as the first two weeks of school for schools that experienced a great deal of new student registration at the beginning of the school year. Any students added to study classrooms after that point were not included in the study.

[76] The overall percentage of students excluded from random assignment was less than 5 percent.

Level 12; fourth graders, the Complete Battery, Level 13; and fifth graders, the Survey Battery, Level 14.

We administered two reading subtests, Reading Comprehension and Vocabulary, and the sum of students' scores represented their total reading scores. For students in grades 2 through 5, we administered two math subtests, Math Concepts and Applications and Math Computation, and the sum of the two scores represented these students' total math scores. However, for kindergarten and first-grade students, no Math Computation subtest exists, so only the Math Concepts and Applications subtest was administered. For comparability across grades, we used only the Math Concepts and Applications subtest as the primary math outcome for all grades in our analyses.

Norm Sample. The instrument was nationally standardized in winter and spring of 1991. The spring standardization involved 115,888 K–12 students drawn from 261 public and 112 private schools; the fall standardization involved 109,825 K–12 students drawn from 265 public and 96 private schools. To ensure that the norm group consisted of a sample representative of the nation's school population, stratified random sampling was used to identify students in the norm groups. Stratification was based on region, community type, school size, and socioeconomic status. The average student in the norm sample had a normal curve equivalent (NCE) score of 50, and the standard deviation (SD) of the NCE scores was 21.06.

Reliability and Validity. Internal consistency (KR20) coefficients ranged from 0.76 to 0.94. There was evidence of content and construct validity.[77]

2. Data on Teachers in the Study

Classroom Performance

We used two methods to collect information on teachers' performance in the classroom: direct observations and interviews with principals. Teachers' performance could be influenced by their training and could also affect the achievement of their students.

Classroom Observations. We conducted direct observations using a version of the Vermont Classroom Observation Tool (VCOT) specially adapted for this study. The VCOT is a proprietary tool for classroom observations developed by the Vermont Institutes over several years. Its precursor was an instrument for measuring the quality of standards-based, investigative science and mathematics instruction, created by Science and Math Program Improvement (SAMPI), a research group at Western Michigan University, and based on research conducted by Horizon Research, Inc. Using the SAMPI Observation Tool as a starting point, Vermont Institutes staff reviewed Charlotte Danielson's Framework for Teaching (1996), on which the widely used Praxis III observational assessment is based

[77] Technical information on the CAT-5 was obtained from "CAT-5 Technical Report," CTB McGraw-Hill, Educational and Professional Publishing Group of The McGraw-Hill Companies, Inc., Monterey, CA, 1996.

(Dwyer 1994). In parallel with the Praxis III content, the VCOT developers included examples of evidence for each indicator, added systematic and ongoing formative and summative assessment of student learning as a major indicator, and simplified and shortened the tool. The VCOT underwent further refinement through its use in the field by a group of trained teacher-leaders who observed classrooms.

In 2004, several of those involved in the original design of the VCOT adapted it for use in the observation of literacy lessons. Development of the literacy version of the VCOT was partly informed by the standards and practices included in the National Council of Teachers of English Standards and the National Reading Panel (NICHD 2000). The VCOT's indicators reflect practices that have been commonly asserted in the literature to be effective (see Danielson 1996), but prior to its use in this study, there had been no research on the relationship between its measures and student achievement. The VCOT is not nationally normed. As for reliability and validity, Cronbach's alpha coefficients for the observations in math and literacy ranged from 0.88 to 0.98 for the sample in this study.

The VCOT covers three domains: lesson implementation, lesson content, and classroom culture.

1. ***Lesson Implementation.*** Indicators in this domain measured the use of best practices, pacing, teacher confidence, and student engagement.

2. ***Lesson Content.*** Indicators in this domain measured the teacher's understanding of the concepts and content of the lesson, applicability of content and class assignments to the real world, and connections to other subjects or lessons.

3. ***Classroom Culture.*** Indicators in this domain measured clarity and consistency of classroom routines, respectfulness and appropriateness of behavior, and teacher sensitivity to student diversity.

The three domains are made up of five, four, and seven indicators, respectively, with each indicator representing a good teaching practice. Examples of the indicators include the following: for lesson implementation, "the pace of the lesson is appropriate for the developmental level of the students"; for literacy lesson content, "understanding of content and concepts is taught through close reading of text and vocabulary instruction"; and for classroom culture, "classroom management maximizes learning opportunities." For each indicator, observers scored teachers as having shown (1) no evidence, (2) limited evidence, (3) moderate evidence, (4) consistent evidence, or (5) extensive evidence. We calculated average scores at the domain level.

VCOT scoring is not on an equal-interval scale. That is, the difference between (1) no evidence and (2) limited evidence may not be the same as between (4) consistent evidence and (5) extensive evidence. However, since the observers were trained to rate the average quality for each domain within 0.75 points of a gold standard panel, we can be confident that we have distinguished between teachers with average scores on the upper and lower portions of the scale (between 2.0 and 4.0, for example). (We cannot necessarily distinguish

between a teacher who scored 2.75 and one who scored 3.5 in the same domain.) In addition, the same observer rated each AC and TC teacher pair, without knowing each teacher's AC or TC classification,, which further ensures that the metric applied to each pair was the same.

To ensure the reliability of ratings of classroom practices, one of the developers of the VCOT trained observers in its use. In both years, the training included instruction, practice sessions observing videotaped lessons, and practice sessions observing lessons in real classrooms. For the 2005–2006 school year, 25 staff members were allowed to conduct observations, but only after meeting the reliability standard set for the evaluation. The potential observer's composite (average) ratings of videotaped lessons in both literacy and math had to fall within 0.75 points of the consensus rating established by a panel of three observers, which included the developer of the VCOT.[78]

Teachers were typically observed on two to four successive days in the spring, teaching a total of two regular literacy (reading and writing) lessons and two regular math lessons. The observations varied in length but lasted an average of one hour for math and 70 minutes for literacy. Teachers knew in advance on which days they would be observed. Giving advance notice was necessary to coordinate the timing of an observation with a literacy or math lesson. Although advance notice may have enabled teachers to perform at a higher-than-average level relative to their typical performance, this would not differentially affect the observations of AC and TC teachers. In cases where original teachers had left, we observed their replacements in 10 of 12 cases.

Principals' Ratings. During our principal interviews in the spring semester of each school year, we asked principals to rate how well each study teacher performed, relative to other teachers in the school, on each of 13 performance indicators, using a 5-point scale where 1 = substantially below average, 3 = average, and 5 = substantially above average. The indicators spanned four domains.

1. ***Reading/Language Arts Instruction.*** Indicators in this domain measured the teacher's ability (1) to discern individual students' learning needs in reading/language arts, (2) to formulate plans to meet those needs, (3) to lead instructional activities, and (4) to modify instruction when necessary to meet individual needs.

2. ***Math Instruction.*** Indicators in this domain measured the same four abilities as for reading/language arts, but with respect to math instruction.

3. ***Classroom Management.*** Indicators in this domain measured how well the teacher (1) establishes and enforces classroom rules and procedures; (2) manages classroom time to keep students on task; (3) enforces desired

[78] Observers could also meet the reliability standard by scoring within 0.75 points of the VCOT developer's rating during a jointly observed lesson in a real classroom.

student behavior through, for example, praise and support; and (4) engages students in learning.

4. ***General.*** One indicator in this domain measured how well the teacher utilizes parents and school resources.

In an effort to reduce potential bias, the forms on which principals marked their ratings did not identify teachers by type of certification program. When original study teachers had left, we asked principals for retrospective ratings of those teachers, as well as current ratings of the replacement teachers whenever possible. In our analysis of these data, we averaged the ratings for each teacher on the four reading/language arts items, the four math items, and the remaining five items. The alpha coefficients from confirmatory factor analyses each exceeded 0.92 for the three constructs.

Background Characteristics

The main source of information on the characteristics of study teachers was a survey administered at the same time the spring achievement tests were given in the classrooms. Information was collected on (1) professional background, including postsecondary institutions and degrees, work history, training programs, and credential status; (2) support (such as mentoring) received during the first year as a full-time teacher; (3) classroom assistance received from a teacher's aide or another teacher; and (4) personal background characteristics, such as age, sex, race/ethnicity, and the number of children they have. We administered the survey to replacement teachers in their classrooms whenever possible. We also administered the survey by mail or phone to original teachers who left before spring testing.

We used additional sources to collect more information on teachers' academic backgrounds and academic achievements. We used *Barron's Profiles of American Colleges 2003* to measure the selectivity of the college or university from which study teachers received their bachelor's degrees.[79] Teachers reported their undergraduate institution in the teacher survey. For both original and replacement teachers who gave us written consent, we also obtained SAT and ACT scores from the respective sponsors of those examinations, the College Board and ACT. We converted ACT scores to SAT scores using concordance tables available from the College Board.[80]

[79] *Barron's* places institutions in six categories of admissions competitiveness, based on admitted freshmen students' high school grades and entrance exam scores, and on the percentage of applicants accepted: (1) most competitive, (2) highly competitive, (3) very competitive, (4) competitive, (5) less competitive, and (6) noncompetitive. In our analyses, we classified teachers whose institutions were in *Barron's* top three categories as having attended a "selective" institution, and those whose institutions were in the next three categories as having attended a "not selective" institution.

[80] Sources: http://www.collegeboard.com/prod_downloads/highered/ra/sat/satACT_concordance.pdf; http://www.collegeboard.com/about/news_info/cbsenior/equiv/rt027027.html.

Certification Program Experiences

Through in-person or telephone interviews with program directors, we collected detailed information on the following aspects of the training programs that original study teachers[81] attended:

- The admission requirements
- The amount, timing, and content of instruction provided
- The amount, nature, and timing of required fieldwork experiences
- The length and features of student teaching assignments for TC teachers
- The nature of any mentoring and support provided to AC teachers during their first year as a teacher of record

All quantitative data collected from these interviews—total hours of instruction and hours in various subject areas (for AC and TC teachers); distribution of total hours before, during, and after first year of teaching (for AC teachers); total hours of fieldwork (for AC and TC teachers); weeks spent in student teaching, hours per day devoted to student teaching, number of full-length school days that candidates were expected to be solely in charge of their classroom during student teaching, number of times student teachers were observed in action by a field supervisor and the average length of these observations, number of times candidates attended a class or seminar associated with student teaching and average length of these sessions, and number of additional meetings with field supervisors and average length of these meetings (for TC teachers)—represent the program directors' best estimates. Interviewers sought written program documents (for example, course lists) and often used worksheets to help prompt and record program directors' best estimates. When possible, we collected information about the unique experiences of the specific teacher in the study sample. When teacher-specific information was not available, we asked questions about these aspects of the programs as they existed when the study teachers were enrolled. Program requirements are a good proxy for the study teachers' actual experiences, because both AC and TC programs were highly prescriptive, requiring candidates to take a set of courses with minimal room for variability within the program.

To make comparisons across these diverse programs, we developed several standard definitions and conventions for describing program characteristics:

- ***Defining the "Program."*** We defined the teacher training *program* as all experiences required for preparing someone to be an elementary school teacher, with a focus on those courses and activities that would provide candidates with contextual and process information on students, classrooms,

[81] We did not seek information on the program experiences of replacement teachers (n=12).

schools, and pedagogy, rather than the content matter they would eventually be teaching. We included courses and fieldwork required after formal admission to the program, as well as education- and teaching-related courses taken (and any affiliated fieldwork) as prerequisites for admission. Courses not directly related to teacher preparation, including non-education-major courses college students take in their first two years in undergraduate-based TC programs, were excluded, as were most undergraduate courses taken by teachers who attended postbaccalaureate TC programs and AC programs; the exception was program prerequisite courses mentioned above.

- ***Defining "Instruction."*** *Instruction* refers to time that candidates are required to spend in class with an instructor in traditional formats such as lectures and seminars—essentially, "seat time" or "contact time"—as well as time that candidates are required to spend completing structured, self-paced instructional assignments, such as computer-based tutorials.[82] We focused on measuring instruction in five areas that, in theory, may be most related to the study's main outcomes. Specifically, we measured instruction in (1) classroom management; (2) reading/language arts pedagogy; (3) math pedagogy; (4) student assessment, defined as how to assess student performance (not how to diagnose learning problems); and (5) child development. Instruction in each of these topics could have been the focus of one or more full courses, or of just part of one or more courses. Any hours of instruction not counted toward one of the five areas of interest was counted as "other" instruction.

- ***Defining "Fieldwork."*** *Fieldwork* refers to time that candidates are required to spend in elementary or middle school classrooms observing teachers and students, working with students, or leading lessons. Student teaching does not count as fieldwork; we defined it as a separate and distinct experience.

- ***Measuring Program Requirements.*** We measured the number of clock hours the candidate spent in various program activities—for example, receiving instruction in certain subjects or doing fieldwork. We converted other metrics that programs commonly use to classify courses, such as credit hours, semester hours, or units, into clock hours after determining how many hours per week a given course or instructional period met and for how many weeks.

- ***Measuring Student Teaching.*** We asked TC program directors about the hours devoted to student teaching each day, the length of the experience in weeks, and the number of full-length school days that student teachers were expected to spend fully in charge of their classrooms, as well as the number of times student teachers were observed by or met with a program staff member, and the average length of these encounters.

[82] This does not include unstructured independent study or time spent preparing for tests or completing course assignments.

C. Response Rates and Student and Teacher Mobility

Exhibit A.1 provides an overview of how students flowed through the study, from random assignment to the point when spring achievement tests were administered. Exhibit A.2 provides an overview of how teachers flowed through the study, including the number of teachers who completed a teacher survey and whose classrooms were observed.

1. Response Rates on Student Data Collection

Data on students came from two sources. First, MPR collected student test score data directly through the administration of the CAT-5, with response rates that ranged from 87 percent to 91 percent for the AC/TC, low-coursework/high-coursework subgroups.[83] Second, each study school provided demographic data found on individual student records. Response rates for these data ranged from 90 percent to 96 percent for the subgroups.

Response rates for data collection on students, shown in Exhibit A.3, were influenced by the availability of school records for student demographics and also by student mobility shown in Exhibit A.4. Approximately 10 percent of study students left the study schools during the study year. Of those, 40 percent moved to another school in the same district and were tested for the study, while the rest left the district and were not tested. Student mobility rates did not differ statistically between AC and TC classrooms.

[83] After students were randomized, we obtained passive consent for them to be part of the study. The passive consent process involved sending letters home with the children and requesting that they be returned only if the parent did not want the child to be part of the study. The response rates in Exhibit A.3 count only those students for whom we had obtained consent. Therefore, the total sample size in Exhibit A.3 differs from the number of students randomized in Exhibit A.1 by the number of students for whom there was no consent.

Exhibit A.1. Flow of Students Through Study

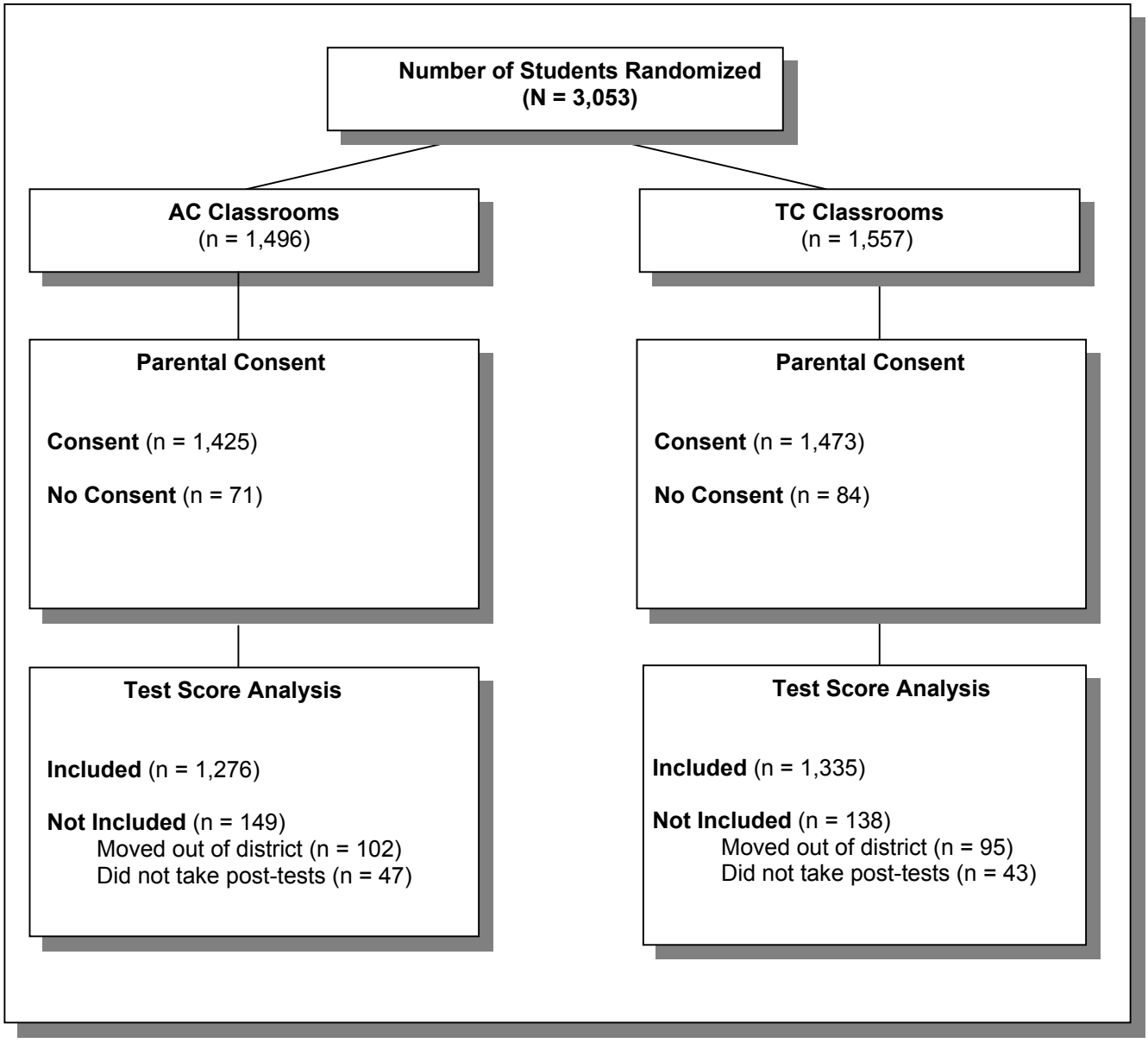
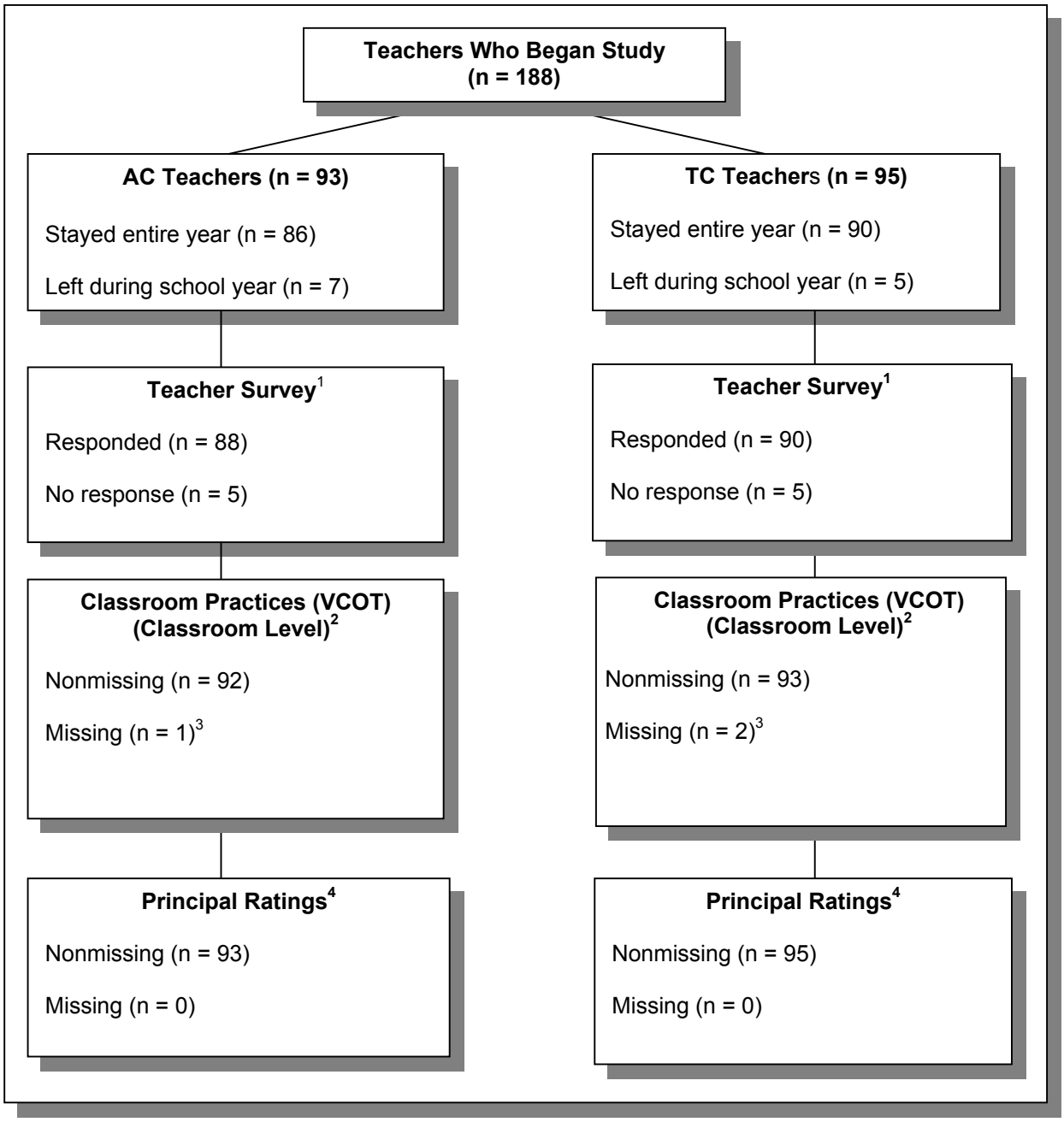


Exhibit A.2. Teachers in the Study



[1] We attempted to survey all teachers, including those who left midyear and those who replaced them. The numbers in this box reflect survey responses for the original study teachers only. We also obtained surveys from three of the teachers replacing AC teachers and four of the teachers replacing TC teachers, though these survey responses were not ultimately used in this study.

[2] The numbers in this box reflect the number of classrooms observed. In cases where the original study teachers left, the replacement teachers were observed. Those instances are counted in these numbers.

[3] The observations for one teacher pair (AC and TC) were lost.

[4] The numbers in this box reflect only the ratings of the original teachers. We also received principal ratings of replacement teachers, though they were not ultimately used in this study.

Exhibit A.3. Response Rates for the Student Sample

	Low-Coursework						High-Coursework					
	AC Classrooms			TC Classrooms			AC Classrooms			TC Classrooms		
	Total	Non-missing	Percentage	Total	Non-missing	Percentage	Total	Non-missing	Percentage	Total	Non-missing	Percentage
Test Scores												
Post-test												
--Reading	754	678	90%	724	653	90%	671	598	89%	749	681	91%
--Math	754	677	90%	724	653	90%	653	582	87%	733	666	89%
Demographics	754	704	93%	724	676	93%						
Gender	754	716	95%	724	694	96%	671	638	95%	749	709	95%
Race/ethnicity	754	704	93%	724	697	96%	671	632	94%	749	707	94%
FRPL[a]	754	678	90%	724	653	90%	671	602	90%	749	672	90%

[a]Free or reduced-price lunch (FRPL) eligibility status was obtained from the same records as the other student demographics. However, some schools refused to release this information, resulting in the lower response rates for this variable.

Exhibit A.4. Mobility of Students in the Sample

	Low Coursework		High Coursework	
	AC Classrooms	TC Classrooms	AC Classrooms	TC Classrooms
Moved within school	67 (8.9%)	44 (6.1%)	40 (6.0%)	49 (6.5%)
Moved out of school, within district	21 (2.8%)	36 (5.0%)	31 (4.6%)	28 (3.7%)
Moved out of district	51 (6.8%)	37 (6.5%)	51 (7.6%)	48 (6.4%)

2. Response Rates on Teacher Data Collections

Data on teachers were gathered from a survey, classroom observations, program interviews, and principal interviews. In the survey, teachers provided information on their undergraduate institution, which we merged to selectivity rankings from *Barron's*. We collected SAT and ACT scores for the teachers who had taken those tests and consented to our request. Response rates were the lowest for the SAT/ACT scores (between 67 and 78 percent) and highest for principal interviews (100 percent).

One factor affecting teacher response rates, shown in Exhibit A.5, was mobility. Twelve original study teachers did not complete the school year in the classroom they were in at the time of random assignment. All the departures took place during the 2005–2006 school year. Most departing teachers (9) left during or at the end of the fall semester, and 10 of the 12 departing teachers were succeeded by a permanent replacement—someone who committed to remain through the end of the school year—rather than by one or more short-term substitute teachers.

Seven of the 12 departing teachers were AC teachers, and 5 were TC teachers. However, all 7 of the departing AC teachers were from low-coursework programs, and 4 of the 5 departing TC teachers were also part of mini-experiments involving low-coursework AC teachers. Thus, comparisons of mini-experiments involving low-coursework AC teachers were more affected by teacher mobility than were comparisons of mini-experiments involving high-coursework AC teachers.

Teacher mobility was concentrated in certain locations. Half the departing teachers were from two schools in District P. In some cases, mobility affected more than one member of a mini-experiment. At one Texas school, the two AC teachers who made up half a quartet at grade 1 were both reassigned early in the fall semester to grade 5 classrooms at the same school. At one New Jersey school, both teachers of a pair left their classrooms.

Exhibit A.5. Response Rates for the Teacher Sample

	Low-Coursework						High-Coursework					
	AC Teachers			TC Teachers			AC Teachers			TC Teachers		
	Total	Non-missing	Percentage	Total	Non-missing	Percentage	Total	Non-missing	Percent	Total	Non-missing	Percentage
Teacher Survey												
Original teachers	51	46	90%	50	46	92%	42	42	100%	45	44	98%
Classroom Observation												
Literacy	51	50	98%	50	49	98%	42	42	100%	45	44	98%
Math	51	50	98%	50	49	98%	41	41	100%	44	43	98%
College Selectivity												
Original teachers	51	40	78%	50	42	84%	42	38	90%	45	42	93%
SAT or ACT Score												
Gave consent	51	49	96%	50	46	92%	42	40	95%	45	42	93%
Took ACT/SAT	51	37	73%	50	43	86%	42	31	74%	45	32	71%
Received scores	51	37	73%	50	40	80%	42	28	67%	45	32	71%
Program director interview	51	50	98%	50	50	100%	42	40	95%	45	44	98%
Principal ratings of classroom performance	51	51	100%	50	50	100%	42	42	100%	45	45	100%

Most departing teachers were replaced by teachers from different kinds of certification programs or with different levels of experience, so students in the departing teachers' classrooms often ended up being taught for at least half the school year by a different type of teacher from the one to whom they had been randomly assigned. Of the 7 departing AC teachers, 2 were replaced by veteran TC teachers, 2 were replaced by novice TC teachers, 1 was replaced by a novice AC teacher, and for 2 we have no information on replacements. Of the 5 departing TC teachers, 2 were replaced by novice TC teachers, 2 were replaced by people who were neither certified nor pursuing certification, and for 1 we have no information on the replacement.

The timing of teachers' departures during the school year affected our ability to collect information from or about them (and their replacements). Three departing teachers and 7 replacements completed the teacher survey. We conducted classroom observations of 2 of the departing teachers and for all 10 of the replacement teachers from other classrooms. Principals provided ratings of the instructional abilities of all departing teachers and 9 of 10 replacements. As mentioned in Chapter II, we did not attempt to conduct program director interviews for any replacement teachers.

The sensitivity of study findings to attrition during the study year is examined in Chapter IV.

D. STATISTICAL POWER

The target sample size for the evaluation was determined during the design phase of the study (Decker et al. 2005). Included below is the table of predicted minimum detectable effect sizes from that design report.

Exhibit A.6. Minimum Effects Under Alternative Sample Designs

Sample	Student Sample Size (Assuming 20 per Class Complete Tests)	Detectable Effect Sizes (Percentage Points) (1) Regression R^2: 60%	(2) Regression R^2: 30%
1. One-Tailed Test. 80 Schools, 180 Teachers, 20 Students Responding per Teacher			
Full Sample	3,600	11	15
50% Subgroup of Programs	1,800	16	21
50% Subgroup of Teachers	1,800	15	20
33% Subgroup of Programs	1,200	20	26
33% Subgroup of Teachers	1,200	19	25
25% Subgroup of Programs	900	23	30
25% Subgroup of Teachers	900	21	28
50% Teachers; 50% Programs	900	22	29
2. Two-Tailed Test. 80 Schools, 180 Teachers, 20 Students Responding per Teacher			
Full Sample	3,600	13	17
50% Subgroup of Programs	1,800	18	24
50% Subgroup of Teachers	1,800	17	23
33% Subgroup of Programs	1,200	22	30
33% Subgroup of Teachers	1,200	21	28
25% Subgroup of Programs	900	26	34
25% Subgroup of Teachers	900	24	32
50% Teachers; 50% Programs	900	25	33

Note: Minimum detectable effects are estimated for a 5 percent level of significance and 80 percent power level. These calculations take into account clustering effects at the teacher level and at the school level. The equation used to calculate the minimum detectable effect is:

$$2.486 * \sqrt{1-R^2} * \sqrt{\frac{2\rho_1(1-c)}{S} + \frac{2\rho_2}{T} + \frac{2(1-\rho_1-\rho_2)}{N}}$$

where

S is the number of schools, *T* is the number of treatment (comparison) teachers, *N* is the number of students in the treatment (comparison) group, *p1* (=0.07) is the between-school variance as a percentage of the total variance of the outcomes based on previous studies, *p2* (=0.16) is the between-teacher variance, and *c* (=0.50) is the correlation between treatment and control group students within the same school. Previous impact evaluations have found that an R^2 as high as 60 percent may be an appropriate assumption when baseline measures of an outcome are available, but 30 percent is more realistic when baseline measures are not available. However, the regression R^2 can vary considerably between studies.

See Decker et al. (2005) for more details.

E. ESTIMATION STRATEGY

1. Experimental Analysis of Student Test Scores

The OLS regression equation used to estimate the achievement effects is:

$$(A.1) \quad y_{ij} = \beta_0 + \beta_1 X_{ij} + \beta_{EXP} EXP_j + \sum_k \beta_k School_k + \sum_k \delta_k School_k AC_j + \varepsilon_{ij}$$

where:

y = student reading or math achievement score (measured in NCEs)
i, j, k = indexes of students, classrooms, and schools
X = student characteristics (baseline test scores in all subjects, race, gender, and free/ reduced-price lunch status)
EXP = a vector of three binary indicators of the years of teaching experience
$School$ = binary indicator of the school the student attended
AC = binary indicator of whether the student was in an AC or a TC classroom
ε = unexplained variation in the outcome
$\beta_0, \beta_1, \beta_{EXP}, \beta_k, \delta_k$ = parameters to be estimated

The school-specific[84] AC effect is represented by δ_k. To calculate the high- and low-coursework AC effects, the δ_k are averaged for each group.[85] The model controls for observable student characteristics that may be correlated with achievement, including the student's race, gender, and fall test scores. Because the AC and TC teachers in our sample vary in their level of experience, we controlled for these differences in the regression. Several studies have noted nonlinear relationships in teacher experience and classroom effectiveness (Hanushek et al. 2005; Boyd et al. 2005), and we allow for nonlinearity by including teacher experience as discrete indicators for one to two years, three to four years, and more than five years.[86] The inclusion of teacher experience does not affect the results; estimates of the model excluding teacher experience are in Exhibit A.8.

[84] For schools with more than one pair of teachers and more than one grade level in the study, this is a school-specific and grade-level-specific effect.

[85] For example, the high-coursework AC effect is calculated as the simple average of the experiment-level AC effects for the 42 mini-experiments involving high-coursework AC teachers. The standard errors are calculated as $\sum_a \sum_b \text{cov}(\delta_a, \delta_b)$ where a and b indicate effects from experiments of the same type of AC teacher (high or low coursework).

[86] The estimates are substantively unchanged when we include a single continuous measure of the teachers' experience instead of the discrete indicators. Appendix Exhibit A.11 shows the estimates that include a continuous measure of experience.

2. Experimental Analysis of Teachers' Classroom Practices

To measure the differences in classroom practices, we estimate:

$$(A.2) \quad P_j = \lambda_0 + \lambda_1 EXP_j + \lambda_2 AC_j + \tau_j$$

where P_j is a measure of teacher practices and EXP_j is a vector of indicators for teacher experience.[87] Teaching experience is modeled as in equation A.1. We do not control for other teacher characteristics, such as measures of academic ability, because these characteristics could be related to the likelihood that a teacher entered the profession through the AC route. The estimates presented in this section are for the combined effect of the type of person who becomes a teacher through the AC route and the training received. Therefore, controlling for teacher characteristics that are correlated with choice of route would obscure the combined effect.

The parameter of primary interest is λ_2, interpreted as the regression-adjusted estimate of the average difference in teacher practices between all AC and all TC teachers. This regression was estimated separately for high-coursework AC and low-coursework AC teachers. Because some mini-experiments had more than one AC or TC teacher, we used weights so that each mini-experiment had equal weight in the overall computation.[88]

Twelve teachers in the study were replaced by other teachers at some point during the year. Although the attrition of teachers was similar for AC and TC teachers as discussed in Chapter II, all the AC teachers who left during the school year were from low-coursework programs, and most of the TC teachers were paired with a low-coursework AC teacher.[89] The potential disruption caused by a teacher departing during the school year should have the largest effect on the intent-to-treat estimate for low-coursework AC teachers.

Findings in the experimental analyses are presented for all the classrooms of the original study teachers, regardless of whether the teacher moved mid-year. When the original study teacher left during the school year, we averaged the effects on observed classroom practices and student outcomes according to the treatment status of the original teacher. For example, if an AC teacher left and was replaced by a TC teacher, the practices of the

[87] For simplicity, we assume that teacher practices are unaffected by the student composition of the classroom. Introducing student characteristics would diminish the precision with which we could estimate effects given the small sample size, but random assignment ensured that the composition of student characteristics in classrooms did not vary systematically within schools.

[88] In four cases, there were three teachers in the experiment, and in one case, there were four. In these cases, the AC impacts for these mini-experiments were calculated as the average AC impact for that school-grade combination.

[89] This suggests that schools or districts in which low-coursework AC teachers were located have higher rates of teacher turnover than places where high-coursework AC teachers were located.

replacement TC teacher were observed but were included in our model with the AC teachers; that is, the classroom was regarded as an AC classroom. Because teacher type might also influence the rate at which students leave classes, the intent-to-treat effects also included the test scores of students who changed classes mid-year and treated their outcomes as if they had remained in their original classes.[90] Because these findings are intended to be internally valid estimates of the teachers in our sample, we treat the teacher effects as fixed and do not correct the standard errors for clustering within classrooms.

3. Nonexperimental Analysis of Differences in Programs, Training, and Teacher Characteristics

We modeled relative teacher effectiveness as a function of the differences in the characteristics of the AC and TC teachers included in each mini-experiment. Specifically, we estimated

$$y_{ijk} = \beta_0 + \beta_x X_{ijk} + \lambda_0 AC_{jk} + \lambda_1 dZ_k AC_{jk} + AC_{jk}\nu_k + \varepsilon_{ijk} \quad \text{(A.3)}$$

where dZ_k is the difference between the AC and TC teachers in school k in some characteristic (such as hours of instruction in a particular subject area or SAT score) and ν_k is an unobserved random variable.

We estimated equation A.3 using ordinary least squares, with clustering accounted for in the standard errors using the Huber-White sandwich estimator. Because our goal was to explain effects with observable differences between the teachers, we restricted the sample to teachers who did not leave the sample during the study. We applied this restriction because it is unclear what relationship the teacher's characteristics had in the overall outcome when the teacher taught for only a portion of the year.

F. SUPPLEMENTARY EXHIBITS FOR CHAPTER IV

On the following pages, we present 11 supplementary tables referenced in Chapter IV.

[90] We were able to follow and complete post-tests only for students who remained in the same district. Over the two cohorts, 200 study students of 2,941 moved out of their districts.

Exhibit A.7. Unadjusted Test Score Differences One Year After Random Assignment in AC and TC Classrooms

	Number of Mini-experiments	AC Teacher Average Score	TC Teacher Average Score	Difference	Effect Size	*p*-Value
Reading						
Overall	90	38.74	38.62	0.12	0.01	0.86
Low coursework	48	38.84	38.50	0.34	0.02	0.72
High coursework	42	38.62	38.75	-0.13	-0.01	0.90
Math						
Overall	89	42.04	42.77	-0.73	-0.03	0.37
Low coursework	48	42.41	42.12	0.30	0.01	0.80
High coursework	41	41.59	43.53	-1.94	-0.09	0.10

Source: California Achievement Test, 5th Edition (CAT-5), administered by Mathematica Policy Research, Inc. (MPR). The math test refers to the Mathematics Concepts subsection. Test scores are expressed in terms of NCEs; the average score nationally is 50, and the SD is 21.06.

Exhibit A.8. Test Score Differences One Year After Random Assignment in AC and TC Classrooms, Omitting Controls for Teacher Experience

	Number of Mini-experiments	AC Teacher Average Score	TC Teacher Average Score	Difference	Effect Size	*p*-Value
Reading						
Overall	90	38.41	38.62	-0.21	-0.01	0.67
Low coursework	42	38.22	38.50	-0.28	-0.01	0.68
High coursework	48	38.62	38.75	-0.13	-0.01	0.86
Math						
Overall	89	41.78	42.77	-0.99	-0.04	0.10
Low coursework	41	41.69	42.12	-0.42	-0.02	0.62
High coursework	48	41.87	43.53	-1.67	-0.07	0.06

Source: California Achievement Test, 5th Edition (CAT-5), administered by Mathematica Policy Research, Inc. (MPR). The math test refers to the Mathematics Concepts subsection. Test scores are expressed in terms of NCEs; the average score nationally is 50, and the SD is 21.06.

Notes: The AC classroom average score reported in the table is the TC average score plus the regression-adjusted treatment effect. The regressions adjust for the students' baseline test scores, eligibility for free or reduced-price lunch, gender, race/ethnicity, and teacher's years of experience as a classroom teacher of record.

Exhibit A.9. Spring Subtest Score Differences in AC and TC Classrooms

	Number of Mini-experiments	AC Teacher Average Score	TC Teacher Average Score	Difference	Effect Size	*p*-Value
Vocabulary						
Overall	90	38.50	39.09	-0.59	-0.03	0.29
Low coursework	48	37.88	38.89	-1.01	-0.05	0.25
High coursework	42	39.21	39.32	-0.11	-0.01	0.89
Comprehension						
Overall	90	38.97	38.65	0.32	0.02	0.62
Low coursework	48	39.21	38.83	0.38	0.02	0.71
High coursework	42	38.70	38.45	0.25	0.01	0.78
Math Computation						
Overall	34	48.06	49.84	-1.78	-0.08	0.26
Low coursework	16	46.28	50.45	-4.17	-0.18	0.04
High coursework	18	49.61	49.26	0.35	0.02	0.90

Source: California Achievement Test, 5th Edition (CAT-5), administered by Mathematica Policy Research, Inc. (MPR). The math computation subtest was administered to students in grades 2-5 only. There were 34 teacher pairs in this sample. Test scores are expressed in terms of NCEs; the average score nationally is 50, and the SD is 21.06.

Notes: The AC classroom average score reported in the table is the TC average score plus the regression-adjusted treatment effect. The regressions adjust for the students' baseline test scores, eligibility for free or reduced-price lunch, gender, race/ethnicity, and teacher's years of experience as a classroom teacher of record.

Exhibit A.10. Spring Test Score Differences in AC and TC Classrooms, Excluding Teachers and Students Who Left During the Study

	Number of Mini-experiments	AC Teacher Average Score	TC Teacher Average Score	Difference	Effect Size	*p*-Value
Low Coursework (N = 916)						
Reading	39	39.30	39.39	-0.09	0.00	0.93
Math	39	49.97	44.58	0.05	0.00	0.95
High Coursework (N = 1,076)						
Reading	42	39.42	39.37	-0.61	0.03	0.63
Math	41	41.62	43.43	–1.81	–0.08	0.05

Source: California Achievement Test, 5th Edition (CAT-5), administered by Mathematica Policy Research, Inc. (MPR). The math test refers to the Mathematics Concepts subsection. Test scores are expressed in terms of NCEs; the average score nationally is 50, and the SD is 21.06.

Notes: These estimates are referred to as the "treatment on the treated" since they include only students and teachers who remained for the entire study. The AC classroom average score reported in the table is the TC average score plus the regression-adjusted treatment effect. The regressions adjust for the students' baseline test scores, eligibility for free or reduced-price lunch, gender, race/ethnicity, and teacher's years of experience as a classroom teacher of record.

Exhibit A.11. Spring Test Score Differences in AC and TC Classrooms, Controlling for Alternative Measures of Teacher Experience

	Number of Mini-experiments	AC Teacher Average Score	TC Teacher Average Score	Difference	Effect Size	*p*-Value
High Coursework						
Certified Experience (Continuous)						
Reading	42	38.65	38.75	–0.10	-0.01	0.89
Math	41	41.96	43.53	–1.57	–0.07	0.09
Instruction Experience (Certified Experience + Long-term Substitute Teaching)						
Reading	42	38.66	38.75	–0.09	0.00	0.90
Math	41	41.96	43.53	–1.58	–0.07	0.08
Total Experience (Instruction Experience + Teacher Aide + Regular Substitute + Misc.)						
Reading	42	39.74	38.75	0.99	0.05	0.23
Math	41	43.10	43.53	–0.43	–0.02	0.66
Low Coursework						
Certified Experience (Continuous)						
Reading	48	38.42	39.50	–0.08	0.00	0.93
Math	48	41.61	42.12	–0.51	–0.02	0.63
Instruction Experience (Certified Experience + Long-term Substitute Teaching)						
Reading	48	38.30	38.50	–0.19	–0.01	0.82
Math	48	41.54	42.12	–0.58	–0.03	0.57
Total Experience (Instruction Experience + Teacher Aide + Regular Substitute + Misc.)						
Reading	48	38.33	38.50	–0.16	–0.01	0.85
Math	48	41.58	42.12	–0.54	–0.02	0.60

Source: California Achievement Test, 5th Edition (CAT-5) and teacher survey, administered by Mathematica Policy Research, Inc. (MPR). The math test refers to the Mathematics Concepts subsection. Test scores are expressed in terms of NCEs; the average score nationally is 50, and the SD is 21.06.

Notes: The benchmark estimation uses discrete indicators for certified experience level. They differ from the estimates in the first rows, which use the continuous measure of this experience. The AC classroom average score reported in the table is the TC average score plus the regression-adjusted treatment effect. The regressions adjust for the students' baseline test scores, eligibility for free or reduced-price lunch, gender, race/ethnicity, and teacher's years of experience as a classroom teacher of record.

Exhibit A.12. Correlations of Within-Pair AC-TC Mean Differences in Program Course Hours and Student Outcomes

	Reading			Math		
	Coeff.	Standard Error	*p*-Value	Coeff.	Standard Error	*p*-Value
AC	-0.86	1.25	0.49	-0.39	1.81	0.83
AC-TC Differences in Required Hours						
Math pedagogy	-0.17	0.30	0.57	-0.52	0.43	0.23
Reading/language arts pedagogy	-0.04	0.13	0.73	0.01	0.18	0.94
Classroom management	0.03	0.19	0.88	-0.10	0.25	0.69
Student assessment	-0.16	0.20	0.43	0.29	0.36	0.43
Child development	0.20	0.17	0.24	-0.10	0.21	0.62
Other instruction	-0.01	0.05	0.83	0.02	0.08	0.76
Fieldwork	-0.02	0.03	0.59	0.05	0.05	0.26
N	**1,921**			**1,895**		

Sources: (1) California Achievement Test, 5th Edition (CAT-5), administered by Mathematica Policy Research, Inc. (MPR). The math test refers to the Mathematics Concepts subsection. Test scores are expressed in terms of NCEs; the average score nationally is 50, and the SD is 21.06. (2) Program director interviews.

Note: The regression model controls for the students' baseline test scores, eligibility for free or reduced-price lunch, gender, race/ethnicity, and teacher's years of experience as a classroom teacher of record.

Exhibit A.13. Correlations of Within-Pair AC-TC Mean Differences in Teacher Training and Student Outcomes

	Reading			Math		
	Coeff.	Standard Error	*p*-Value	Coeff.	Standard Error	*p*-Value
AC	0.07	1.07	0.95	-0.35	1.39	0.80
AC-TC Differences Teacher Training/Preparation						
Master's degree	-2.44	1.03	0.02	-1.87	1.29	0.15
Education major	0.19	0.93	0.84	-0.78	1.49	0.60
Business/Math major	1.45	1.31	0.27	-2.37	1.80	0.19
Had mentor, first year	0.75	1.01	0.46	-1.02	1.82	0.57
Had regular communication with supervisor, first year	-0.05	0.97	0.96	-1.65	1.37	0.23
Had chance to observe classes, first year	-0.29	1.02	0.78	-1.07	1.53	0.49
Currently taking courses	-2.53	0.92	0.01	-1.14	1.34	0.39
N	**1,959**			**1,933**		

Sources: (1) California Achievement Test, 5th Edition (CAT-5), administered by Mathematica Policy Research, Inc. (MPR). The math test refers to the Mathematics Concepts subsection. Test scores are expressed in terms of NCEs; the average score nationally is 50, and the SD is 21.06. (2) Teacher survey.

Note: The regression model controls for the students' baseline test scores, eligibility for free or reduced-price lunch, gender, race/ethnicity, and teacher's years of experience as a classroom teacher of record.

Exhibit A.14. Correlations of Within-Pair AC-TC Mean Differences in Teacher Characteristics and Student Outcomes

	Reading			Math		
	Coeff.	Standard Error	*p*-Value	Coeff.	Standard Error	*p*-Value
AC	-0.47	1.00	0.64	-0.40	1.47	0.79
AC-TC Differences in Teacher Characteristics						
Experience	-0.58	0.45	0.20	-1.12	0.63	0.08
Selective college	-0.34	1.06	0.75	-0.38	1.62	0.82
SAT score	-0.01	0.00	0.11	-0.01	0.01	0.23
Black	0.37	1.41	0.80	-1.35	1.99	0.50
Hispanic/Latino	0.66	0.97	0.50	-2.09	2.39	0.38
Female	-0.54	1.29	0.68	-0.23	3.14	0.94
Age	0.02	0.08	0.78	0.00	0.11	0.98
N	**1,746**			**1,744**		

Sources: (1) California Achievement Test, 5th Edition (CAT-5), administered by Mathematica Policy Research, Inc. (MPR). The math test refers to the Mathematics Concepts subsection. Test scores are expressed in terms of NCEs; the average score nationally is 50, and the SD is 21.06. (2) Teacher survey. (3) *Barron's Profiles of American Colleges*. (4) The College Board and ACT.

Notes: The regression model controls for the students' baseline test scores, eligibility for free or reduced-price lunch, gender, race/ethnicity, and teacher's years of experience as a classroom teacher of record. *Selective college* is defined as a college or university rated by *Barron's* as in the top three competitiveness categories: most competitive, highly competitive, or very competitive. SAT score includes the SAT equivalent of an ACT score, where necessary.

Exhibit A.15. Interactions of Students' and Teachers' Race/Ethnicity

	Reading			Math		
	Coeff.	Standard Error	*p*-Value	Coeff.	Standard Error	*p*-Value
Black student has black teacher	2.35	1.72	0.17	3.45	2.03	0.09
Hispanic student has Hispanic teacher	–2.52	2.97	0.40	–1.43	2.52	0.57
N	**1,959**			**1,991**		

Sources: (1) California Achievement Test, 5th Edition (CAT-5), administered by Mathematica Policy Research, Inc. (MPR). The math test refers to the Mathematics Concepts subsection. Test scores are expressed in terms of NCEs; the average score nationally is 50, and the SD is 21.06. (2) Teacher survey. (3) School records.

Notes: These estimates are restricted to the "treatment on the treated" sample of teachers and students who did not leave during the study. Coefficients are from ordinary least squares models that include experiment-level fixed effects to control for all unobserved differences across schools. This model is specified as

$$y_{ijk} = \beta_0 + \lambda_0 AC_{jk} + \lambda_1 BTea * BStu_{ijk} + \lambda_2 HTea * HStu_{ijk} + \lambda_3 BStu_{ijk} + \lambda_4 HStu_{ijk} + \lambda_5 BTea_{jk} + \lambda_6 HTea_{jk} + \beta_x X_{ijk} + \nu_j + \varepsilon_{ijk}$$

where *BTea*BStu* is dummy variable indicating that a student is a black student in the class of a black teacher, *HTea*HStu* is a dummy indicating that the student is a Hispanic student in the class of a Hispanic teacher, *BTea* is an indicator that the teacher is black, *HTea* is an indicator that the teacher is Hispanic, *BStu* is an indicator that the student is black, *HStu* is an indicator that the student is Hispanic, *X* is a vector of other student characteristics (baseline test scores in all subjects, race, gender, and free/reduced-price lunch status), and *i, j,k* index student, class, and pair. The coefficient λ_1 is the marginal effect for a black student assigned to the class of a black teacher and λ_2 is the marginal effect for a Hispanic student assigned to the class of a Hispanic teacher.

Exhibit A.16. Correlations of Within-Pair AC-TC Mean Differences in Teaching Practices and Student Outcomes

	Reading			Math		
	Coeff.	Standard Error	*p*-Value	Coeff.	Standard Error	*p*-Value
AC	-0.60	0.86	0.49	-1.20	1.09	0.27
AC-TC Differences Teacher Practice Ratings						
Literacy implementation	1.15	2.09	0.58	2.18	2.73	0.43
Literacy culture	-3.67	1.44	0.01	-2.30	2.04	0.26
Literacy content	-0.04	2.22	0.99	-0.13	3.14	0.97
Math implementation	-0.49	2.05	0.81	1.33	2.39	0.58
Math culture	1.72	1.47	0.25	0.40	1.91	0.84
Math content	1.37	2.15	0.53	1.29	2.99	0.67
N	**1,965**			**1,964**		

Sources: (1) California Achievement Test, 5th Edition (CAT-5), administered by Mathematica Policy Research, Inc. (MPR). The math test refers to the Mathematics Concepts subsection. Test scores are expressed in terms of NCEs; the average score nationally is 50, and the SD is 21.06. (2) Ratings on the VCOT.

Note: The regression model controls for the students' baseline test scores, eligibility for free or reduced-price lunch, gender, race/ethnicity, and teacher's years of experience as a classroom teacher of record.

Exhibit A.17. Correlations of Within-Pair AC-TC Mean Differences in Principal Ratings and Student Outcomes

	Reading			Math		
	Coeff.	Standard Error	*p*-Value	Coeff.	Standard Error	*p*-Value
AC	-0.04	0.87	0.96	-1.24	1.08	0.25
AC-TC Differences in Principal Ratings						
Reading/Language Arts	3.05	1.69	0.07	3.50	2.41	0.15
Math	-0.61	1.55	0.70	-0.61	1.96	0.76
Classroom management	-2.13	0.93	0.02	-1.67	1.14	0.15
N	**1,991**			**1,965**		

Sources: (1) California Achievement Test, 5th Edition (CAT-5), administered by Mathematica Policy Research, Inc. (MPR). The math test refers to the Mathematics Concepts subsection. Test scores are expressed in terms of NCEs; the average score nationally is 50, and the SD is 21.06. (2) Principal interviews.

Note: The regression model controls for the students' baseline test scores, eligibility for free or reduced-price lunch, gender, race/ethnicity, and teacher's years of experience.