NCEE 2013-4002

U. S. DEPARTMENT OF EDUCATION

Effects of the Pacific CHILD Professional Development Program

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

Effects of the Pacific CHILD Professional Development Program

Final Report

December 2012

Authors:

Yasuyo Abe
Berkeley Policy Associates

Vanora Thomas
Berkeley Policy Associates

Castle Sinicrope
Social Policy Research Associates

Kevin A. Gee
Brown University

Project Officer:
Karen Armstrong
Institute of Education Sciences

NCEE 2013–4002
U.S. Department of Education

Arne Duncan

Secretary

Institute of Education Sciences

John Q. Easton

Director

National Center for Education Evaluation and Regional Assistance

Ruth Curran Neild

Commissioner

December 2012

This report was prepared for the National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, under contract ED-06-CO-0024 with Regional Educational Laboratory Pacific administered by Berkeley Policy Associates.

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 report.

This report is in the public domain. Authorization to reproduce it in whole or in part is granted. While permission to reprint this publication is not necessary, the citation should read: Abe, Y., Thomas, V., Sinicrope, C., & Gee, K. A. (2012). *Effects of the Pacific CHILD Professional Development Program*. (NCEE 2013–4002). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

This report is available on the Institute of Education Sciences website at http://ncee.ed.gov and the Regional Educational Laboratory Program website at http://edlabs.ed.gov.

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

Disclosure of potential conflict of interest

The research team for this evaluation is Berkeley Policy Associates (BPA) of Oakland, California. Neither Berkeley Policy Associates nor any of its key staff members have financial interests that could be affected by findings from the evaluation of Pacific CHILD. Throughout the study, the BPA research team staff, as a third-party evaluator, worked independently of and separately from the REL Pacific staff who developed and implemented the Pacific CHILD. Recruitment was the one activity that required collaboration by the two teams. During recruitment, the BPA research team provided potential participants with technical information on the research design and requirements for data collection activities. It also supported REL Pacific staff on the guidelines for selecting sample schools with respect to the target number of schools to be recruited and the criteria for prioritizing the schools to target. No one in the Technical Working Group, convened by the BPA research team to provide advice and guidance, has financial interests that could be affected by findings from the evaluation.

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.

Acknowledgments

The authors would like to acknowledge colleagues and partners who made the study possible from the early design phases to the final analysis. This study represents a collaborative effort of state education agencies, schools, principals, teachers, Regional Educational Laboratory Pacific (REL Pacific), and researchers. We appreciate the willingness of the principals, schools, and teachers to respond to many requests for data and access to classrooms.

We would like to thank the REL Pacific's *Pacific Communities with High Performance in Literacy Development* (Pacific CHILD) developers and professional development team, including Marylin Low, Margaret Ho, Jennifer Padua, and the professional development providers who brought Pacific CHILD to schools in the three study entities. The Pacific CHILD Reading Advisory Panel provided REL Pacific with valuable design and development guidance as well as feedback on Pacific CHILD professional development program. Panel members included Anne Cunningham, University of California, Berkeley; Gerald Duffy, University of North Carolina, Greensboro; Jana Echevarria, California State University, Long Beach; Michael Kamil, Chairman, Stanford University; and Dorothy Strickland, Rutgers University. We were also fortunate to have the advice of an expert technical working group. Members included Geoffrey Borman, University of Wisconsin–Madison; Daniel Brown, University of Hawai‘i at Hilo; Margo Gottlieb, Illinois Resource Center; Rosa Salas Palomo, University of Guam; Hirokazu Yoshikawa, Harvard University; and Shuqiang Zhang, University of Hawai‘i at Mānoa. We also thank anonymous reviewers who provided valuable feedback, through the Institute of Education Sciences, on this report. We would also like to thank the individuals at state education agencies (SEA) who provided the research team and REL Pacific with support from recruitment through the final stages of data collection. SEA staff include Claire Poumele, American Samoa Department of Education (ASDOE) director; Philo Jennings, ASDOE Deputy Director of the Department of Education; Donna Gurr, Assistant Director of the ASDOE's Office of Curriculum, Instruction and Accountability; Elizabeth Haleck Filiaga, Assistant Director for the ASDOE Student Information System and Services Division; Rita Sablan, Commissioner, the Commonwealth of Northern Mariana Islands Public School System (CNMI PSS); Jackie Quitugua, Associate Commissioner for Instructional Services, CNMI PSS; Fasefulu Tigilau, Assessment Coordinator, CNMI PSS; Pat Hammamoto, former Superintendent, Hawai‘i Department of Education (HIDOE); Cara Tanimura, Director of the HIDOE Systems Accountability Office; Kent Hinton and Jerald Plett, HIDOE Systems Accountability Office.

Special thanks go to Roger Chesswas at REL Pacific for his leadership and continued support for this study and to Johannes Bos at the American Institutes for Research (AIR) for his earlier leadership in the study at BPA, continued guidance on navigating the review process, and his advice and feedback. Without their consistent support throughout the study years, this research would not have been possible. We also thank the staff at Pacific Resources for Education and Learning and Berkeley Policy Associates who contributed to and assisted in all aspects of this research study and report, including Don Burger, Denise Perdue, Zee Oshiro, Linda Toms Barker, Raquel Sanchez, Paola Rochabrun Oliveira, Kristin Bard, Naomi Tyler, and Dina de Veer. Finally, we would like to pay tribute to the memory of Thomas Barlow, a lifetime educator who dedicated his work to improving the well-being of children in the Pacific and provided tireless support to this study as Director of REL Pacific.

Contents

DISCLOSURE OF POTENTIAL CONFLICT OF INTEREST ..... **III**

ACKNOWLEDGMENTS ..... **V**

EXECUTIVE SUMMARY ..... **XIII**

Study design xiv
Outcome measures xiv
Study sample xv
Implementation of Pacific CHILD xvi
Analytic methods and impact findings xvi
Subgroup and exploratory analyses xvii
Study limitations xviii

CHAPTER 1: INTRODUCTION AND STUDY OVERVIEW **1**

What the literature says about professional development in reading 1
Study context 2
Description of Pacific CHILD 5
Overview of the study design 8
Content and organization of this report 10

CHAPTER 2: STUDY DESIGN **11**

Random assignment design 11
School recruitment and selection 12
Random assignment of schools 15
Student and teacher samples for the impact analysis 20
Analysis framework 31

CHAPTER 3: DATA **37**

Outcome measures and instruments 37
School, student, and teacher background data 42
Implementation data 43
Data collection rates, response rates, and quality assurance 45

CHAPTER 4: IMPLEMENTATION OF PACIFIC CHILD **49**

Description of Pacific CHILD activities as designed 49
Actual exposure to Pacific CHILD 51
Actual delivery of Pacific CHILD activities 59
Summary 61

CHAPTER 5: IMPACT ANALYSIS FINDINGS **63**

Impact of Pacific CHILD on students 63
Impact of Pacific CHILD on teachers 65

CHAPTER 6: EXPLORATORY ANALYSES **69**

Impact of Pacific CHILD on students and teachers in Hawai'i 69
Impact of Pacific CHILD on students in American Samoa and the Commonwealth of the Northern Mariana Islands 73
Impact of Pacific CHILD on subscales of teacher practice 74
Impact of moderating factors on teacher impacts 76

CHAPTER 7: SUMMARY OF STUDY DESIGN, CONFIRMATORY FINDINGS, AND STUDY LIMITATIONS 79

Review of study design and implementation of Pacific CHILD 79
Effect of Pacific CHILD on student achievement in reading comprehension 80
Effect of Pacific CHILD on teacher knowledge and practice 80
Study limitations and caveats.....81

APPENDIX A: STATISTICAL POWER ANALYSIS A1

APPENDIX B: TECHNICAL NOTES ON IMPACT ANALYSES..... B1

Analytic approach B1
Impact estimation results B9
Sensitivity analyses..... B12
Analyses of additional teacher outcome measures B17

APPENDIX C: TECHNICAL NOTES ON EXPLORATORY ANALYSES AND RESULTS OF ANALYSIS OF IMPACT ON TEACHERS IN NON-HAWAIʻI SUBSAMPLES C1

Analytic approach C1
Analysis of Hawaiʻi and non-Hawaiʻi subsamples..... C3
Analyses of subscales of teacher practice measure C9
Analyses of moderating effects of teacher characteristics C10
Analyses of impact on teachers in American Samoa and the Commonwealth of the Northern Mariana Islands..... C12

APPENDIX D: TEACHER OUTCOME MEASURES D1

Measuring teacher knowledge: The teacher knowledge assessment..... D1
Measuring teacher practice: The teacher practice observation protocol D3
Limitations to the teacher outcome measures..... D5

APPENDIX E: STUDENT OUTCOME MEASURES E1

Student achievement data E1
Properties of standardized tests used to measure student reading comprehension E1
Combining student achievement data E2
Limitations to the student outcome measure E3

APPENDIX F: ADAPTATIONS TO DESIGN AND CONTEXT OF IMPLEMENTATION.....F1

Adaptations to the intervention.....F1
Context of implementation F4

APPENDIX G: EXPOSURE TO PACIFIC CHILD BY STUDENTS IN THE IMPACT SAMPLE G1

APPENDIX H: EXPOSURE TO PACIFIC CHILD BY PARTICIPATING TEACHERS IN THE IMPACT SAMPLE H1

APPENDIX I: EXPOSURE TO PACIFIC CHILD BY TEACHERS IN IMPACT SAMPLE WHO WERE PRESENT AT TREATMENT GROUP SCHOOLS DURING ENTIRE INTERVENTION PERIOD I1

APPENDIX J: TEACHERS' EXPOSURE TO PACIFIC CHILD BY ENTITY..... J1

APPENDIX K: QUALITATIVE METHODS USED TO ANALYZE ADAPTATIONS TO DESIGN AND CONTEXT OF IMPLEMENTATION K1

Data sources and preparation K2
Training K4
Qualitative analysis K4
Cross-verification of data through triangulation K6

APPENDIX L: INSTRUMENTS L1

L.1 Teacher knowledge assessment L2
L.2 Additional eight items for the modified SIOP L10
L.3 Teacher background survey L14
L.4 Principal survey: Treatment group schools L30
L.5 Principal survey: Control group schools L36
L.6 Pacific CHILD interview guide: Program developer L41
L.7 Pacific CHILD interview guide: Project advisors L46
L.8 Pacific CHILD interview guide: Trainers L51
L.9 Teacher focus group discussion guide L55
L.10 Professional development observation summary for Annual Institute Week 1 ..... L60
L.11 Professional development observation summary for Annual Institute Week 2 ..... L63
L.12 Professional development observation summary for Mini Institute full day L66
L.13 Professional development observation summary for Mini Institute school-based day L69
L.14 Program staff lesson demonstration observation summary L71
L.15 Teacher lesson demonstration observation summary L72
L.16 Structured Learning Team observation (SLT) L73

REFERENCES R1

Figures

Figure 1.1 Map of the Pacific region 3
Figure 1.2 Theoretical model of Pacific CHILD 9
Figure 2.1 Movement of teachers at study schools 22
Figure 2.2 Flow chart of sample selection, from random assignment of schools to selection of impact analysis sample 24

Tables

Table 2.1 Characteristics of recruited schools 14
Table 2.2 Characteristics of schools in impact sample, by entity 17
Table 2.3 Baseline characteristics for treatment and control group schools 19
Table 2.4 Movement of students at study schools 23
Table 2.5 Profile of student analysis sample, by entity 27
Table 2.6 Profile of teacher analysis sample, by entity 28
Table 2.7 Characteristics of students at treatment and control group schools 29
Table 2.8 Characteristics of teachers at treatment and control group schools 30
Table 3.1 Data collection schedule for outcome, background, and implementation data 39
Table 3.2 Summary of data types and purposes 45
Table 3.3 Completion and response rates for students and teachers for data collected to measure impact of Pacific CHILD 46
Table 3.4 Comparison of data completion and response rates between the treatment and control group schools 47
Table 4.1 Sequence of intervention activities and components 50
Table 4.2 Prescribed annual exposure to Pacific CHILD 53
Table 4.3 Attendance at Pacific CHILD annual institutes by teachers at treatment group schools in impact sample 54
Table 4.4 Attendance at Pacific CHILD mini-institutes by teachers at treatment group schools in impact sample 55
Table 4.5 Year-round Pacific CHILD support received per month by teachers at treatment group schools in impact sample 56
Table 4.6 Summary of exposure to Pacific CHILD by teachers in impact sample 57
Table 4.7 Exposure to Pacific CHILD by teachers at treatment group schools in impact sample, by level of participation 58
Table 4.8 Prescribed and actual duration of year-round Pacific CHILD support activities delivered to teachers at treatment group schools in impact sample over two-year intervention 59
Table 5.1 Scale scores on reading comprehension component of Stanford 10 Achievement Test of students at control group schools 63
Table 5.2 Intent-to-treat impact estimates of Stanford 10 Achievement Test scale score for reading comprehension 64
Table 5.3 Outcome measures for teachers at control group schools at baseline 66
Table 5.4 Intent-to-treat impact estimates on teacher outcomes 67
Table 6.1 Estimated impact of Pacific CHILD on Stanford 10 Achievement Test–equivalent reading comprehension scale scores of grade 5 students in Hawai'i 71
Table 6.2 Estimated impact of Pacific CHILD on knowledge and practice of teachers in Hawai'i 72
Table 6.3 Estimated impact of Pacific CHILD on reading comprehension of grade 5 students in American Samoa 74
Table 6.4 Estimated impact of Pacific CHILD on reading comprehension of grade 5 students in the Commonwealth of the Northern Mariana Islands 74
Table 6.5 Estimated impact of Pacific CHILD on teacher practice subscales with Cronbach's alpha equal to or greater than .70 76
Table 6.6 Estimated impact of teacher experience and education on impact of Pacific CHILD on teacher outcomes 78

Table A.1 Power analysis based on original assumptions about sample size for measuring impact of Pacific CHILD in study schools A2
Table A.2 Power analysis based on revised assumptions about sample size for measuring impact of Pacific CHILD in study schools A2
Table B.1 Preliminary specification tests: *p*-values from likelihood ratio test B3
Table B.2 Covariates used to estimate impact of Pacific CHILD in confirmatory analyses B6
Table B.3 Summary of outcome variables and covariates used to estimate impact of Pacific CHILD in confirmatory analyses B8
Table B.4 Unadjusted means of outcome measures in treatment and control schools B10
Table B.5 Estimated impacts of Pacific CHILD on grade 5 Stanford 10 Achievement Test reading comprehension scale scores B11
Table B.6 Sensitivity analyses for estimated impact of Pacific CHILD on student outcomes B14
Table B.7 Sensitivity analyses of impact of Pacific CHILD on teacher knowledge assessment scores based on alternative estimation methods B15
Table B.8 Sensitivity analyses of impact of Pacific CHILD on teacher practice observation scores based on alternative estimation methods B16
Table B.9 Estimated impact of Pacific CHILD on teacher knowledge and practice based on alternative outcome measures B18
Table C.1 Minimum detectable effect size for analysis of exploratory questions on impact of Pacific CHILD on students and teachers C3
Table C.2 Profiles of student samples used to assess impact of Pacific CHILD C4
Table C.3 Covariates included in exploratory analyses of impact of Pacific CHILD in Hawai'i and non-Hawai'i student and teachers subsamples C5
Table C.4 Estimated impact of Pacific CHILD on reading comprehension of grade 5 students in Hawai'i , alternative covariates C8
Table C.5 Summary of subscale measures of teacher practice measure C9
Table C.6 Estimated impact of Pacific CHILD on teacher practice subscales with Cronbach's alpha less than .70 ..... C10
Table C.7 Effect of additional year of experience or advanced degree on estimated impact of Pacific CHILD on Hawai'i teacher outcomes C11
Table C.8 Estimates of moderating effects of additional year of teaching experience on impact on total score of teacher knowledge assessment C12
Table C.9 Estimated impact of Pacific CHILD on knowledge and practice of grade 4 and grade 5 teachers in American Samoa and the Commonwealth of the Northern Mariana Islands (combined) C13
Table E.1 Length, average difficulty, and reliability of assessments used to measure reading comprehension in grade 5 students, by entity E2
Table E.2 Scale scores, linked scale scores, and *z*-scores for measures of student achievement in reading comprehension E3
Table F.1 Adaptations to the component delivery schedule F3
Table H.1 Prescribed and actual attendance at Pacific CHILD annual institute by participating teachers at treatment group schools in impact sample H1
Table H.2 Prescribed and actual attendance at Pacific CHILD mini-institutes by participating teachers at treatment group schools in impact sample H2
Table H.3 Prescribed and actual year-round Pacific CHILD support per month for participating teachers at treatment group schools in impact sample H3
Table H.4 Prescribed and actual exposure to Pacific CHILD by participating teachers at treatment group schools in impact sample H3
Table I.1 Prescribed and actual attendance at Pacific CHILD annual institute by teachers in impact sample who were present at treatment group schools during entire intervention period I1
Table I.2 Prescribed and actual attendance at Pacific CHILD mini-institutes by teachers in impact sample who were present at treatment group schools during entire intervention period I2
Table I.3 Prescribed and actual year-round Pacific CHILD support per month of teachers in impact sample present at treatment group schools during entire intervention period I3

TABLE I.4 PRESCRIBED AND ACTUAL EXPOSURE TO PACIFIC CHILD TEACHERS IN IMPACT SAMPLE WHO WERE PRESENT AT TREATMENT GROUP SCHOOLS DURING ENTIRE INTERVENTION PERIOD I3
TABLE J.1 TOTAL EXPOSURE TO PACIFIC CHILD BY TEACHERS AT TREATMENT GROUP SCHOOLS IN IMPACT SAMPLE, BY ENTITY J1
TABLE K.1 INDICATORS FOR DETERMINING FIDELITY OF IMPLEMENTATION TO PACIFIC CHILD DESIGN K2
TABLE K.2 SAMPLE OF PRE-ESTABLISHED CODES USED TO CATEGORIZE PACIFIC CHILD ANNUAL INSTITUTE AND MINI-INSTITUTE OBSERVATIONS K5
TABLE K.3 SOURCES OF DATA FOR ASSESSING FIDELITY OF PACIFIC CHILD IMPLEMENTATION K7

Executive Summary

Major federal education initiatives, including the No Child Left Behind Act of 2001, have highlighted the importance of teacher quality in improving student achievement. The federal government has committed significant funding and resources to professional development programs for teachers through Title II of the Act and other initiatives. There is limited evidence, however, of the effectiveness of professional development programs in improving teacher knowledge, teacher practice, and student achievement in reading. This report contributes to the body of research on professional development by presenting the results of an impact study of the Pacific Communities with High Performance in Literacy Development (Pacific CHILD), a professional development program designed and implemented by the Regional Educational Laboratory Pacific (REL Pacific).

REL Pacific staff designed the Pacific CHILD program to improve the instruction of reading comprehension by grade 4 and grade 5 teachers and to raise student achievement in reading comprehension in the Pacific Region. REL Pacific studied an earlier version of the program using an observational design (Chesswas et al. 2005). This study is the first rigorous test of the effectiveness of the program.

REL Pacific staff implemented Pacific CHILD in elementary schools in three entities in the Pacific region: American Samoa, the Commonwealth of the Northern Mariana Islands (CNMI), and Hawai‘i. Pacific CHILD is designed to be a two-year school-based intervention that provides sustained, year-round instructional support to grade 4 and grade 5 teachers in English language arts. The content of Pacific CHILD consists of six components that combine three strategies for improving student reading comprehension (vocabulary, text structure, and question generation) with three strategies for improving classroom pedagogy (differentiated instruction, cognitively rich environments, and interactive tasks).

Pacific CHILD is designed to provide 42 days of professional development over the course of two years. Each year of the intervention consists of the following activities: one 10-day annual institute, three 3-day mini-institutes (one full day and two half-days), monthly lesson demonstrations, twice-monthly classroom observations, and weekly meetings of structured learning teams. The annual institute and mini-institutes consist of workshop-style professional development and hands-on practice in classrooms with students. During the monthly demonstrations, program staff model exemplary practices while teachers observe, question, and reflect on the lesson demonstration. During the twice-monthly classroom observations, program staff observe teachers implementing Pacific CHILD components in their own classrooms. Weekly school-based structured learning teams serve as collaborative learning communities and are designed to facilitate dialogue among teachers and program staff about their experiences with Pacific CHILD. Staff lead the structured learning team meetings twice a month; during the other meetings, teachers meet on their own, guided by an agenda.

Study design

The study was designed to assess whether Pacific CHILD improved student achievement in reading comprehension (primary outcome) and teacher knowledge and practice (secondary outcomes) in the three entities in which it was implemented. It was guided by the following research questions:

Primary research question regarding impacts on student reading comprehension:

- Did grade 5 students at schools that were offered Pacific CHILD for two years perform differently on assessments of reading comprehension from grade 5 students at schools that were not offered Pacific CHILD?

Secondary research question regarding impacts on teacher knowledge and teacher practice:

- Did grade 4 and grade 5 teachers at schools that were offered Pacific CHILD for two years perform differently from teachers at schools that were not offered Pacific CHILD, as measured by either an assessment of their knowledge of theories and strategies related to effective reading instruction (including English language learner-focused theories and strategies) or an assessment of their instructional practices for enhancing student reading comprehension (including English language learner-focused practices)?

To investigate these questions, the study conducted a multisite, cluster randomized experiment in which schools were the unit of random assignment and teachers and students at schools were the units of analysis. The treatment condition consisted of offering schools the opportunity for their grade 4 and grade 5 teachers to participate in the two-year Pacific CHILD professional development program. The control condition consisted of business as usual; schools in the control group were not offered Pacific CHILD during the two years while the treatment group schools were offered the intervention. The study estimated the intent-to-treat effects of Pacific CHILD as a school-level intervention, focusing on individuals who could have been potentially exposed to the full two-year intervention.

Outcome measures

Student achievement in reading was measured using the reading comprehension subtests of the national, norm-referenced tests each entity administers as part of its regular student assessment (the Stanford 10 Achievement Test [SAT 10] in American Samoa and the CNMI and the TerraNova, 2nd Edition, in Hawai‘i). For the impact analyses, the study team converted TerraNova reading comprehension scale scores from Hawai‘i into estimated SAT 10 reading comprehension equivalents, using published norming tables and equipercentile methods.

Impacts on teacher knowledge were measured with a written teacher knowledge assessment developed for this study. Impacts on teacher practice were measured using a modified version of the Sheltered Instruction Observation Protocol (SIOP®) (Echevarria, Vogt, and Short 2007), expanded for this study with items relevant to Pacific CHILD.

Study sample

The study was based on a convenience sample of public elementary schools in American Samoa, the CNMI, and Hawai‘i. In American Samoa and the CNMI, recruitment was completed during the 2006/07 school year, and the intervention started in the summer of 2007. In Hawai‘i recruitment efforts continued into the 2007/08 school year, and the intervention started in the summer of 2008. Fifty-one schools were originally recruited to participate in the study and were randomly assigned to treatment or control conditions. The schools were blocked to ensure that the resulting allocation of schools in the treatment and control groups was balanced in terms of both the number of schools and key school characteristics. After random assignment of schools, teachers in participating schools were recruited to participate in the study.

The study estimated the intent-to-treat effects of Pacific CHILD as a school-level intervention. The study therefore did not establish a fixed sample of students or teachers at the time of random assignment to be tracked through the study period. Instead, the student impact sample consisted of all grade 5 students enrolled in the study schools at the time of data collection in the spring of the second year of the intervention. The teacher impact sample consisted of all grade 4 and grade 5 teachers who were teaching English language arts in a self-contained classroom at the study schools at the time of data collection in the spring of the second year of the intervention. The individual samples for impact analyses were thus defined after random assignment. The impact estimates based on these samples thus represented the effects on individuals at the schools that were offered Pacific CHILD.

During the first year of the program implementation, the study team received a report that, prior to the first annual institute, school administrators in some treatment schools had replaced teachers who were not able to attend the institute with teachers from another grade. These treatment schools, along with other schools in the same assignment blocks, were removed from the sample used to conduct the impact analyses because of the possibility that the reassignment decisions were systematically carried out in direct response to the study treatment itself and the potential compromise that this posed to the integrity of the experimental design. In total, six schools (three treatment and three control schools) were removed. Thus final study sample for the impact analyses consisted of 45 schools, of which 23 were assigned to the treatment condition and 22 to the control condition.

.The analysis sample for measuring impacts on achievement in reading comprehension consisted of 3,052 students, with 1,566 in the treatment group and 1,486 in the control group. For the student sample, missing outcome data were deemed minimal (with the data completion rate of 99.2 percent overall) and balanced across conditions (with the data completion rate of 98.7 percent in the treatment group and 99.7 percent in the control group).

The analysis sample for measuring the impact on teacher knowledge included 197 teachers, with 95 in the treatment group and 102 in the control group. The data completion rate for the teacher knowledge assessment was thus 83.5 percent overall, with 80.5 percent in the treatment group and 86.4 percent in the control group. The analysis sample for measuring the impact on teacher practice included 198 teachers, with 96 in the treatment group and 102 in the control group. The data completion rate for teacher classroom observations was 83.9 percent overall, with 81.4 percent in the treatment group and 86.4 percent in the control group.

The extent of data completion among teachers reflects both their consent to participate in the collection of their outcome data as well as their actual response to the outcome data collection efforts. The data completion rate is thus computed as a product of the consent rate among all teachers in the impact sample and the response rate among the consented. The study found that the difference in data completion rates between the treatment and control group teachers was largely reflective of the difference in the their consent rates; the consent rate among the impact sample teachers was 81.4 percent in treatment schools and 88.1 percent in control schools.

Implementation of Pacific CHILD

Teachers at treatment group schools in the impact sample were not exposed to Pacific CHILD at the levels prescribed by the intervention. Over the course of two years of the intervention, the 118 teachers in the treatment group, including the 50 teachers who did not participate in any program activities, were exposed to an average of 15 of the 42 days prescribed (36 percent of the prescribed intervention). Teachers who participated in the program during both the first and second years of the intervention averaged 31 days of exposure (74 percent of the prescribed intervention). The difference between the prescribed levels of participation and the average actual level of participation largely reflects the fact that 42 percent of teachers in the impact treatment group sample did not participate in any Pacific CHILD program activities during the two-year program and were therefore not exposed to any of the intervention.

Fidelity to the original intervention delivery design varied across program activities. During the two-year intervention, the annual institute and mini-institutes were implemented with the prescribed frequency; however, mini-institutes were not consistently delivered over the course of three consecutive calendar days, as designed. Year-round activities were not implemented as frequently as prescribed, but the average duration of the year-round activities met or exceeded the designed duration.

Analytic methods and impact findings

The impact analysis used a hierarchical linear model as the primary statistical model for estimating the impacts of Pacific CHILD. The study specified a two-level model, in which teachers and students were nested within schools to account for the effects of clustering of individuals within each school. The minimum detectable effect size for the study was estimated a priori to be 0.16 for student outcomes and 0.46 for teacher outcomes. The models were estimated using a restricted maximum likelihood method. They included covariates for school and individual background characteristics and for assignment blocks. Missing data were handled using listwise deletion. Sensitivity analyses were conducted to test whether the findings remained robust to model specifications and estimation methods.

Overall impact was estimated by pooling estimates for the three entities. For the student impact analysis, impact was computed as the weighted mean of the impacts estimated within each entity. For the teacher impact analysis, the impact was estimated based on a sample that pooled observations across all three entities.

The primary impact analysis investigated whether Pacific CHILD improved students' achievement in reading comprehension, as measured by the SAT 10. Based on the benchmark

model estimation, the study finds a statistically significant difference between the treatment and control groups in SAT 10 reading comprehension scores: the estimated average score was 634.3 for students at treatment schools and 629.0 for students at control schools (effect size = 0.244, p = .017).

The secondary impact analysis investigated whether Pacific CHILD improved teacher knowledge or teacher practice. After adjusting p-values for multiple testing based on the Benjamini-Hochberg procedure, the study finds statistically significant differences between treatment and control group schools for both teacher outcome measures. For teacher knowledge, the estimated average score on a 40-point knowledge assessment was 27.0 points for teachers at treatment group schools and 25.0 points for teachers at control group schools. This difference is statistically significant (effect size = 0.35, adjusted p = .023). For teacher practice, the estimated average observation score (on a five-point scale ranging from zero to four) was 2.20 in the treatment group and 1.85 in the control group. This difference is statistically significant (effect size = 0.64, adjusted p = .006). Alternative estimation methods yielded consistent results, supporting the conclusion that Pacific CHILD had impacts across both teacher outcomes.

Subgroup and exploratory analyses

To supplement the impact results and generate potential hypotheses for future investigation, the study explored the patterns of impact on (a) students and teachers in the Hawai‘i subsample, (b) students in the non-Hawai‘i subsamples, (c) subscales of the teacher practice measure, and (d) subgroups of teachers. The same analytic methods used in the confirmatory impact analyses were used in the exploratory analyses. Highlights of the findings from the exploratory analyses include the following:

- Student effects varied across entity. In the Hawai‘i and CNMI subsamples, differences between the reading comprehension scores of treatment and control group students are statistically significant (effect size = 0.10, p = .037 in Hawai‘i ; effect size = 0.36, p = .025 in the CNMI). In the American Samoa subsample, the impact is not statistically significant (effect size = –0.15, p = .629). These results suggest that the effectiveness of Pacific CHILD varied across the study entities.
- In the Hawai‘i subsample, the difference in classroom practice ratings between teachers at treatment and control group schools is statistically significant (effect size = 0.66, p = .018). Differences in teacher knowledge are not statistically significant (effect size = 0.11, p = .640).
- Differences in classroom practice between teachers at treatment and control group schools for the four subscales of the observation protocol for which Cronbach’s alphas were 0.70 or higher are statistically significant. These subscales included two of the three main dimensions: preparation (effect size = 0.50, p = .013) and instruction (effect size = 0.51, p = .016). Impacts for two subareas of instruction: strategies (effect size = 0.47, p = .037) and lesson delivery (effect size = 0.41, p = .024) are also statistically significant.
- More experienced teachers scored higher on the teacher knowledge assessment, with an additional year of experience associated with a difference in impact of 0.21 points (p = .023). On average, the impact estimate for teachers with five years of teaching experience was 1.4

points (effect size = 0.25), whereas the impact for teachers with six years of experience was 1.6 points (effect size = 0.29). The number of years of teaching experience did not have a statistically significant moderating effect on the impact of the intervention for teacher practice observation ratings (p = .912). Having completed an advanced degree did not have a statistically significant moderating effect on the impact of the intervention for either teacher knowledge (p = .787) or teacher practice (p = .793).

Study limitations

The study has limitations that should be considered when reviewing the results. They include, but are not limited to, the limited generalizability of the findings, the composition of the American Samoa sample, the validity of outcome measures, sample equivalence, and sample attrition.

Limited ability to generalize results

Given that the three entities—American Samoa, the CNMI, and Hawai‘i —were purposefully selected as the study sites, the findings in this report are not generalizable to the broader Pacific region. Furthermore, because the study schools in each entity were a convenience sample, the findings are not representative even of the entities themselves. Thus, inferences about the impacts of Pacific CHILD cannot be generalized beyond the study schools.

Although the underlying theoretical model of Pacific CHILD was based on assumptions about individual-level responses to the intervention, the study did not directly address the question of whether Pacific CHILD had an impact on individuals who were offered the intervention. Instead, it examined the effects on individuals in schools that were offered the intervention. For this reason, findings from this study are not intended to support conclusions about the intent-to-treat effects on individuals offered the intervention.

Composition of the American Samoa sample

In American Samoa, the exclusion of schools suspected of having compromised the integrity of the experimental design led to the removal of larger schools from the original sample recruited for this study. The study sample in American Samoa thus did not represent the full range of school types (sizes) targeted by the original sample design.

Validity of outcome measures

There could be concerns that the national, norm-referenced tests, such as those used to measure student outcomes in this study, are not appropriate for measuring achievement in reading comprehension among English language learners and within a diverse regional cultural context. The use of the national, norm-referenced tests was justified for this study, despite the large number of English language learners involved and the particular regional context, on the grounds that they were widely used measures of reading achievement and regarded as policy-relevant tools.

To measure teacher knowledge and classroom practice, the study team developed one instrument and adapted another from an existing tool. Use of instruments developed or adapted for this study may raise concerns about overalignment with the intervention and the validity of the inferences based on data collected using the instruments. In addition, for the teacher practice measure, bias may arise due to the observers' knowledge of the assignment condition of schools. Steps were taken to address these concerns and ensure that the instruments support valid inferences about impacts.

Sample equivalence and sample attrition

Systematic differences in the impact analysis sample between treatment and control conditions could lead to potential bias in the impact estimates. This was of particular concern for teachers in this study. Given that the impact teacher sample was a cohort defined toward the end of the two-year intervention, factors such as teachers' knowledge of their schools' assignment status and their own exposure to the intervention to date—i.e., factors that differed across the conditions—could have influenced their participation in the outcome data collection. If treatment group teachers who selected to participate in the outcome data collection efforts differed both in observed and unobserved ways from control group teachers who selected to participate, the assumption of sample equivalence in expectation across the conditions could be compromised, leading to potential bias in the study results

One indicator for determining whether the equivalence between control and treatment conditions was compromised is sample attrition. In this study, sample attrition for the student and teacher samples reflects the extent to which outcome data collection was not completed. For the student impact sample, as noted above, the rate of outcome data completion was high and consistent across the conditions. For the student impact sample, the overall attrition rate was 0.8 percent, with a differential attrition rate between the treatment and control groups of 1.0 percent. For the teacher impact sample, the overall attrition rate for the teacher impact sample—which includes those teachers who did not consent to participate in the study as well as those who consented but did not respond to the data collection—was 16 percent for both the teacher knowledge assessment and teacher practice observations. The differential attrition rate was 6 percent for the teacher knowledge assessment and 5 percent for the teacher practice observations

Based on the attrition bias model and the bias threshold supported by the What Works Clearinghouse guidelines, the combination of the overall and differential attrition rates resulted in an acceptable level of bias for the student outcome data. According to the same guidelines, the attrition rates resulted in an acceptable level for the teacher practice observation data. For the teacher knowledge data, it resulted in a potentially acceptable level of bias data, suggesting that readers should consider potential attrition bias in evaluating the teacher impact analysis results.

Chapter 1: Introduction and study overview

Major federal education initiatives, including the No Child Left Behind Act of 2001, have highlighted the importance of teacher quality in improving student achievement and committed significant funding and resources to professional development of teachers. Although Title II of the act and other federal programs continue to provide millions of dollars in funding for teacher professional development programs each year (Birman et al. 2007; U.S. Department of Education 2005), there is only limited evidence of the effectiveness of professional development programs in improving teacher knowledge, teacher practice, and student achievement in reading comprehension. This report contributes to the body of research on professional development in reading by presenting the results of an impact study of a professional development program designed and implemented by the Regional Educational Laboratory Pacific (REL Pacific).

REL Pacific staff designed the Pacific Communities with High Performance in Literacy Development (Pacific CHILD) program to improve the reading instruction of grade 4 and grade 5 teachers in order to raise student reading comprehension achievement in the Pacific region. Pacific CHILD is a two-year, year-round professional development intervention that combines workshops with lesson demonstrations, classroom observations, and peer learning groups. REL Pacific staff implemented Pacific CHILD between 2007 and 2010 in elementary schools in three entities in the Pacific region: American Samoa, the Commonwealth of the Northern Mariana Islands (CNMI), and Hawai‘i.[1] An independent research team from Berkeley Policy Associates conducted a rigorous study of the impacts of the program on student achievement in reading comprehension, teacher knowledge, and teacher practice.[2]

This chapter provides background information on and an overview of the study. It opens by reviewing the literature on the impact of teacher professional development on reading and English language arts. It then describes the study context and the intervention, including its research base, content, and structure. It concludes with a discussion of the theoretical model for the intervention and a brief overview of the study design and research questions.

What the literature says about professional development in reading

Despite decades of studies on teacher professional development, limited evidence exists to guide practitioners, policymakers, and researchers in selecting programs focused on reading (Wayne et al. 2008). Of the many studies of teacher professional development in reading, few use randomized control trials or quasi-experimental designs—the kinds of rigorous designs that support causal inferences about effectiveness. For example, a 2007 review of more than 1,300 studies on teacher professional development and student achievement in a range of academic

[1] Pacific CHILD was implemented in American Samoa and the CNMI during the 2007/08 and 2008/09 school years. It was implemented one year later in Hawai‘i, during the 2008/09 and 2009/10 school years.

[2] Berkeley Policy Associates, under a subcontract to REL Pacific, led the study, with funding from the Institute of Education Sciences, the U.S. Department of Education. The study consists of REL Pacific’s implementation of Pacific CHILD and Berkeley Policy Associates’ independent evaluation. This report was prepared by the Berkeley Policy Associates research team.

subjects identifies only six rigorous studies that examine the relationship between professional development and student achievement in reading and English language arts (Yoon et al. 2007).

Studies with rigorous designs find inconsistent evidence of the effects of teacher professional development on teacher knowledge, teacher practice, and student achievement in reading and English language arts.[3] Of the six studies identified by Yoon et al., three show positive and statistically significant impacts in some or all areas of student achievement (Cole 1992; McGill-Franzen et al. 1999; McCutchen et al. 2002).[4] The other three studies show no impacts on student achievement (Duffy et al. 1986; Sloan 1993; Tienken 2003). Although studies published after the review by Yoon et al. find positive and statistically significant impacts for professional development programs on teacher knowledge and teacher practice, these impacts have not translated into gains in student achievement in reading (Garet et al. 2008; Gersten, Dimino, and Jayanthi 2010).

Given the small number of rigorous studies of professional development programs in reading and English language arts, it is not possible to draw conclusions about features of effective programs. The studies cited above suggest that the length and intensity of professional development may be related to its effectiveness. Studies of programs that provided 5–20 contact hours per teacher over two to nine months do not find statistically significant impacts on student achievement (Duffy et al. 1986; Sloan 1993; Tienken 2003; Gersten et al. 2010), although one study (Gersten et al. 2010) finds impacts on teacher knowledge and teacher practice. Three of the five studies of the more intensive professional development programs, including programs that provided 30–100 contact hours per teacher over four months to a year, find positive and statistically significant impacts on student achievement in reading (Cole 1992; McCutchen et al. 2002; McGill-Franzen et al. 1999). One study of two intensive professional development programs that consisted of 48 and 60 contact hours of professional development finds positive and statistically significant impacts on teacher knowledge and some teacher practices but not on student achievement in reading (Garet et al. 2008).

Study context

This study was designed to assess the effectiveness of the Pacific CHILD professional development program, a two-year, year-round intervention designed to improve student reading achievement, teacher knowledge, and teacher practice. The expected number of hours of professional development in the two-year Pacific CHILD program was 295 contact hours per teacher. While REL Pacific studied an earlier version of the Pacific CHILD program using an observational design (Chesswas et al. 2005), the current study is the first rigorous test of the effectiveness of the Pacific CHILD professional development program.

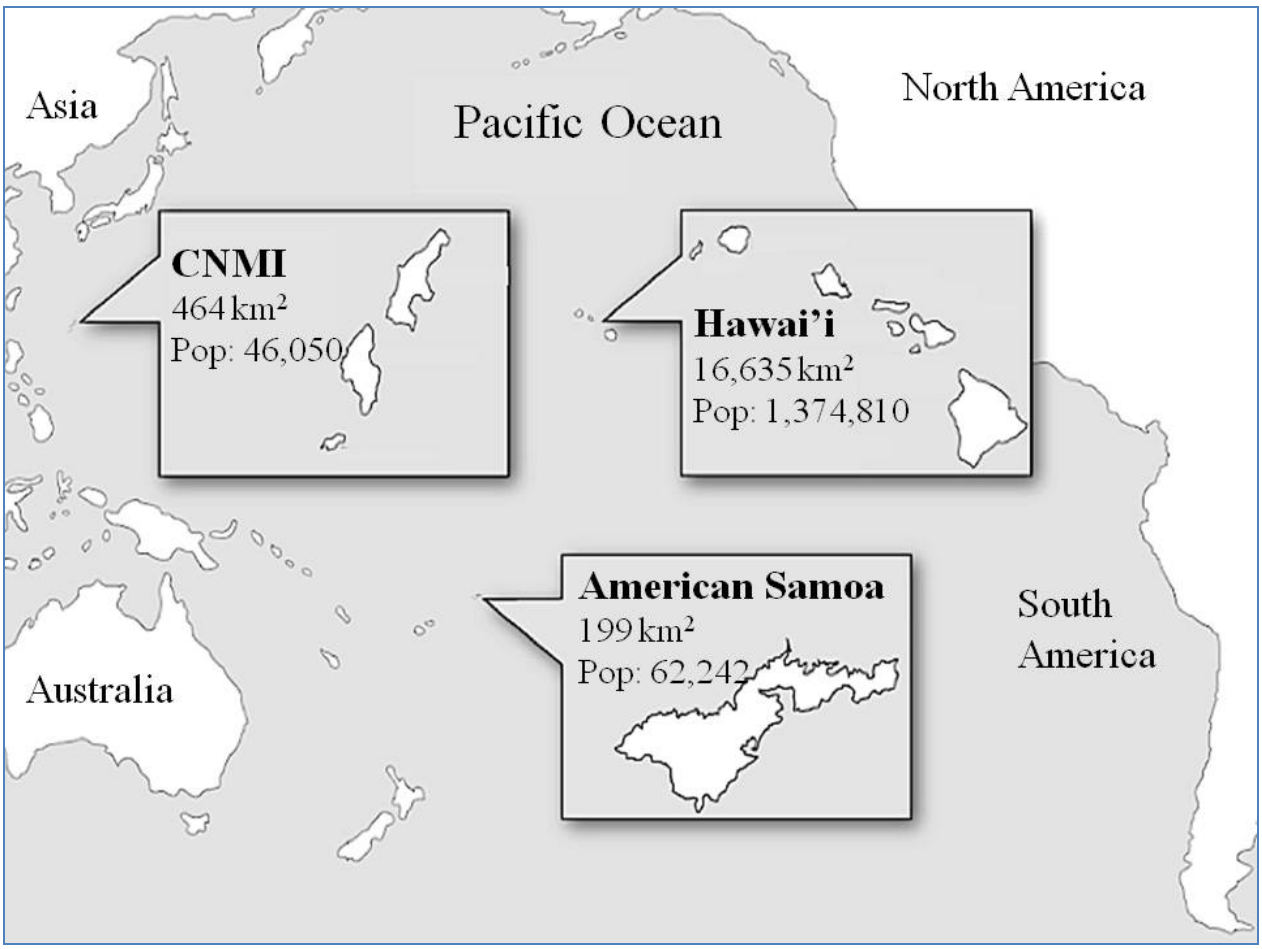
The study estimates the impacts of the Pacific CHILD professional development program in public elementary schools in American Samoa, the CNMI, and Hawai‘i (figure 1.1). These entities represent a convenience sample of sites in the Pacific region; they were selected based on administrative support from their respective education agencies and the availability of student

[3] The studies reviewed focus on student outcomes in reading/ language arts rather than solely on reading achievement.

[4] Two of the six studies that focus on teacher professional development and student achievement in reading and English language arts also focus on student achievement in other areas: Cole (1992) focuses on mathematics, reading, and language; Sloan (1993) focuses on reading, mathematics, and science.

achievement measures that could be analyzed across the entities. The following sections provide background information on student reading achievement and teacher preparedness in the three entities.

Figure 1.1 Map of the Pacific region



Source: Population and area data are from Central Intelligence Agency (2011) and U.S. Census Bureau (2011a, 2011b).

Student reading achievement

Elementary school students in all three entities in the study performed below the national average on standardized tests of reading achievement in 2007. In Hawai‘i, 25 percent of grade 4 students scored at or above proficient on the reading component of the National Assessment of Educational Progress (NAEP), 6 percentage points below the national average of 31 percent (Lee, Grigg, and Donahue 2007). American Samoa and the CNMI do not participate in the NAEP, but both administer the Stanford 10 Achievement Test (SAT 10), a nationally normed standardized test, to select grades.[5] In 2007 students in American Samoa and the CNMI

[5] The NAEP was administered to grade 8 students in American Samoa in 2002; just 1 percent of students scored at or above proficient (Grigg et al. 2003). No other student cohorts in American Samoa or the CNMI have participated in the NAEP since 1992. Although American Samoa and the CNMI are not required by law to comply with the No Child Left

performed below average on reading achievement on the SAT 10.[6] The average grade 4 student in American Samoa scored at the 13th national percentile in reading on the SAT 10; the average grade 5 student in the CNMI scored at the 37th percentile (American Samoa Department of Education 2007b; CNMI Public School System 2007). Students in American Samoa and the CNMI often do not receive sustained exposure to English until they begin formal schooling, and more than 80 percent of them are reported to be English language learners (Burger, Mauricio, and Ryan 2007).

Teacher preparedness

The education level and professional preparedness of teachers vary across the three entities. In American Samoa, teachers holding less than an associate's degree with no formal training in education can receive a temporary certification and begin teaching in the classroom (American Samoa Department of Education 2007a).[7] In 2003, 40 percent of teachers in American Samoa held at least a bachelor's degree, and 8 percent held advanced degrees (Heine and Emesiochl 2007).

Guidelines from the CNMI Public School System state that teachers must hold bachelor's degrees and pass general and content area exams to receive basic certification (CNMI Public School System n.d.). [8] In 2003, 99 percent of teachers in the CNMI held at least a bachelor's degree, and 17 percent held advanced degrees (Heine and Emesiochl 2007).

In Hawai‘i, prospective teachers must complete a state-approved teacher education program from an accredited institution and pass both general and content-area exams to become licensed teachers (Hawai‘i Teacher Standards Board n.d.).[9] In 2007 all teachers in Hawai‘i held at least a bachelor's degree, and 29 percent held advanced degrees (Hawai‘i Department of Education 2007).

The percentage of teachers working on waivers without full state certification or licensure provides another perspective on teacher preparedness in the study entities. In 2006 the share of teachers on waivers was 63 percent in American Samoa, 17 percent in the CNMI, and 4 percent in Hawai‘i (U.S. Department of Education 2009). In addition to having lower education levels and lower licensure rates than teachers in Hawai‘i, nearly all teachers in American Samoa and the CNMI speak English as a second (or third) language, with varying levels of proficiency (Burger, Mauricio, and Ryan 2007; Hunkin-Finau 2007).

Behind Act of 2001, both entities have developed statewide assessments based on their state-level standards. Implementation of the state-level, standards-based assessments began in 2008.

[6] Percentile ranks for the SAT 10 are based on the sample used to norm the tests during development.

[7] In 2008 American Samoa passed the Teacher Reclassification Bill, which limits the length of time teachers can hold temporary and provisional certification and to incentivize teacher certification. The bill instituted a revised pay scale for teachers to reward professional certification and status as "highly qualified" with increased wages. It also requires that teachers with temporary or provisional certifications enroll in approved and accredited programs to earn an associate's or bachelor's degree.

[8] Documentation from the CNMI Public School System states that teachers must pass the Praxis I and II exams, administered by the Educational Testing Services.

[9] The Hawai‘i Teacher Standards Board requires teachers to pass the Praxis I and the Praxis II: Elementary Education: Content Knowledge, to become licensed to teach elementary school in Hawai‘i. As needed, the Hawai‘i Department of Education hires teachers in the process of becoming licensed.

Description of Pacific CHILD

The Pacific CHILD professional development program is a two-year, year-round professional development program for teachers of grade 4 and grade 5 English language arts. The program was designed to improve instruction in reading comprehension in the Pacific region, with a focus on meeting the needs of English language learner students.[10] Pacific CHILD was designed to be applicable across the Pacific region by supplementing rather than replacing existing English language arts curricula, using local staff to deliver the program, and being flexible enough to meet the cultural and linguistic needs of teachers and students in the Pacific region.[11]

Research-based content

The content of the Pacific CHILD professional development program was developed by REL Pacific based on research on the hypothesized effectiveness of reading and instructional strategies.[12] In particular, REL Pacific staff drew on findings presented in the 2000 report of the National Reading Panel on the role of vocabulary instruction, question generation, and other reading strategies in building reading skills. REL Pacific also drew on findings from research on instructional effectiveness in the classroom, particularly for English language learner students, that suggests that (a) teachers should differentiate their instruction based on the diverse needs of students, (b) cognitively-rich environments stimulate literacy development, and (c) interactive tasks promote student learning in the classroom.[13] During the development of Pacific CHILD, REL Pacific staff consulted with its reading advisory panel of national experts.

The Pacific CHILD professional development program consists of six components, which combine three reading strategies with three instructional strategies. Its three reading strategies include the following:

- *Vocabulary*: Vocabulary focuses on meaning at the word level. It is a key element of reading comprehension. Pacific CHILD encourages teachers to use strategies to explicitly teach vocabulary words and word parts (root words, prefixes, suffixes) to help students solve unknown words and to increase students' vocabularies. To enhance students' vocabulary, Pacific CHILD also emphasizes the use of expository texts that are rich in academic words.[14]
- *Text structure*: Text structure refers to the way in which writers organize or structure the ideas in their writing to communicate their message to readers. Knowledge of text structures

---

[10] Although Pacific CHILD focuses on the needs of English language learners, it was designed to improve the reading achievement of all students in grades 4 and 5. The program was designed to be implemented in mainstream classrooms with both English language learners and students proficient in English.

[11] All REL Pacific staff have taught in the local school system to which they are assigned. They help teachers select texts that are culturally and linguistically appropriate for teachers and students in the local setting.

[12] The Pacific CHILD developers cite the following research as the basis for their professional development model: Center for Research on Education, Diversity, and Excellence (2002); Chesswas et al. (2005); Darling-Hammond and McLaughlin (1995); Fullan (2001); Joyce and Showers (1996); Learning First Alliance (2000); Rueda (1998); Strickland and Kamil (2004); Tharp (1997); Tharp et al. (2000).

[13] The Pacific CHILD developers cite research on instructional effectiveness from Echevarria, Vogt, and Short (2003); Genesee (1994); Gibbons (2006); and Johnson and Johnson (1999).

[14] Expository text is nonfiction reading material that provides information on, explains, or describes a subject. Textbooks, trade books, guides, and articles include expository text. Program developers expected that expository text would be a focus of grade 4 and 5 reading instruction.

is expected to improve students' reading comprehension by helping them understand the purpose of the text, anticipate where information will be found in the text, and make predictions about the content. Text features are the physical features of text, such as headings, captions, and illustrations, that are commonly used in expository text. Pacific CHILD focuses on the text structures of compare and contrast and cause and effect as well as on the use of text features.

- *Question generation*: Question generation refers to activities that (a) involve asking and answering questions based on the text and (b) require students to understand what is written, synthesize information, and make generalizations. Question generation is expected to improve students' reading comprehension by focusing their tasks and activities on finding the main idea or important ideas, asking questions, and answering questions about texts. Pacific CHILD provides teachers with strategies that emphasize question generation and use of questioning by students in the classroom.

Its three instructional strategies including the following:

- *Differentiated instruction*: Differentiated instruction is an educational approach in which teachers adjust their instruction based on student needs. Students are expected to learn more effectively in classrooms in which teachers differentiate content, processes, and completed products based on student needs and ability levels. Pacific CHILD promotes differentiated instruction and encourages teachers to take student language ability, reading ability, interests, and readiness into consideration when planning lessons.
- *Cognitively rich environment*: Cognitively rich environments engage students in learning by immersing them in print- and literacy-rich opportunities and experiential learning. Cognitively rich environments are expected to stimulate the language, visual, spatial, emotional, and kinesthetic regions of the brain to engage students, promote learning, and allow students to understand how learning expands beyond the classroom. Pacific CHLD provides teachers with examples of cognitively rich environments and strategies for creating such environments.[15]
- *Interactive tasks*: Interactive tasks are activities that promote student-to-teacher and student-to-student language interaction. Such tasks provide opportunities for students to work in pairs, in small groups, and in large groups. Within groups, students are assigned tasks that promote their use of language and require them to work together to understand and apply ideas, learn from shared experiences, solve problems, and think critically about issues. Through this shared interaction, students are expected to learn from one another and the teacher and expand their knowledge base. Pacific CHILD focuses on interactive tasks and encourages teachers to adopt group-based activities in the classroom.

Program structure and activities

REL Pacific developed the structure and activities of the Pacific CHILD professional development program based on six principles drawn from previous research on adult learning, change, diversity, and effective professional development practices for building reading

[15] Examples of cognitively rich environments include visual representations such as graphic organizers, charts, diagrams, and maps; student- and teacher-generated text; and classroom libraries with a variety of narrative and information-rich texts.

instruction skills.[16] These principles suggest that teachers learn best when professional development providers:

- Design a variety of experiences that address broad program goals as well as real challenges encountered by students and teachers in the classroom.
- Base their ongoing support for teachers on observations of, and appropriate feedback on, actual practice.
- Promote situations in which teachers share their expertise in professional dialogue.
- Challenge teachers to find the best, rather than the easiest, solutions to challenges.
- Engage teachers in collaborative conversations.
- Design activities that take the cultural and linguistic diversity of teachers into account.

The Pacific CHILD program includes lesson demonstrations and observations in the classroom, year-round instructional support and feedback from program staff, and collaboration and professional dialogue among teachers in peer learning groups.[17] The program includes the following activities each year of the two-year program:

- *One 10-day annual institute*: The 10-day annual institute consists of five days of off-site, workshop-style professional development and five days of practice opportunities in classrooms with students. [18] The annual institute takes place during the summer.
- *Three 3-day mini-institutes*: Each mini-institute consists of one day of lecture and small group work and two half days of in-school demonstration lessons, classroom observations, and debriefing sessions led by program staff. The mini-institutes take place during the school year.[19]
- *Monthly lesson demonstrations*: Each month, program staff model exemplary practices in the classroom while teachers observe, question, and reflect on the demonstration.
- *Twice-monthly classroom observations*: Twice a month, program staff observe teacher lessons. Each classroom observation consists of a preconference, a classroom observation, and a postconference.
- *Weekly structured learning team meetings*: Every other week, program staff lead structured learning team meetings. On alternate weeks, participating teachers facilitate their own meetings.

---

[16] The developers cite the following research as the basis for their professional development model: Center for Research on Education, Diversity, and Excellence (2002); Chesswas et al. (2005); Darling-Hammond and McLaughlin (1995); Fullan (2001); Joyce and Showers (1996); Learning First Alliance (2000); Rueda (1998); Strickland and Kamil (2004); Tharp (1997); Tharp et al. (2000).

[17] Throughout the rest of the chapter, the term *program staff* is used to refer to the REL Pacific staff who implemented Pacific CHILD in participating schools.

[18] Students from nonstudy schools were recruited for practice instructional opportunities during the second week of the annual institutes. Annual institutes were held in June, July, or August before each study year. The exact dates were determined based on program staff and teacher availability.

[19] Mini-institutes were held during the following months: mini-institutes 1 and 4: September–October; mini-institute 2 and 5: November–January; mini-institutes 3 and 6: February–May

Program staff provide the program content—the three reading strategy components and the three instructional strategy components—to teachers through these activities using a spiral approach, in which teachers revisit the content repeatedly and in greater depth over the course of the two years.[20] Chapter 4 provides additional information on the structure and activities of the intervention over the two-year period.

Overview of the study design

This study was designed to assess whether Pacific CHILD improved teacher knowledge, teacher practice, and student achievement. The following sections provide a brief overview of the study design, including a discussion of the theoretical model of Pacific CHILD, a brief introduction to the research questions and design, and a roadmap to the rest of the report.

Theoretical model of Pacific CHILD

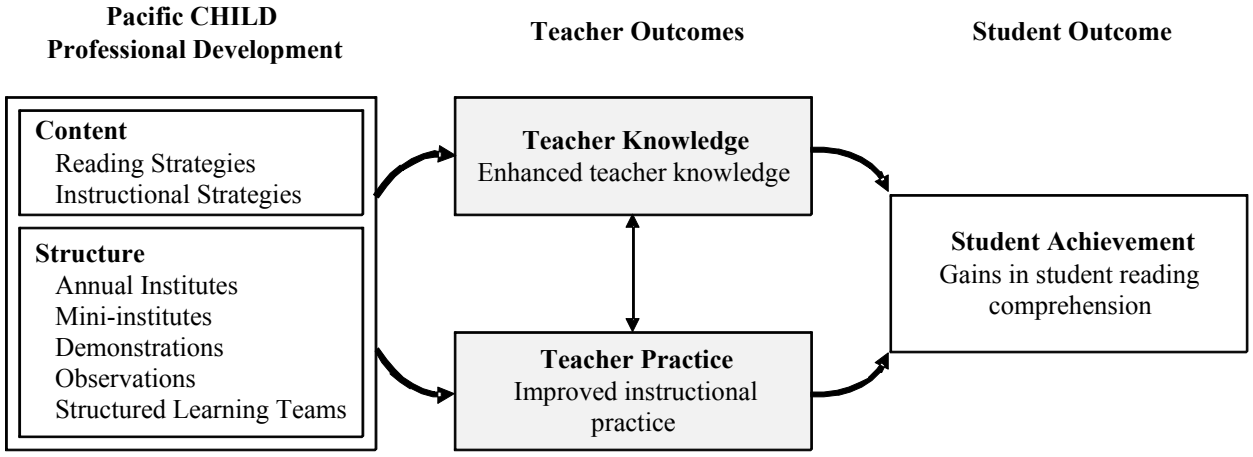
This study is guided by a theoretical model that describes the relationship between the program and its intended impacts on student achievement in reading comprehension, teacher knowledge, and teacher practice. The model is illustrated in figure 1.2. The arrows in the figure represent the relationships between the intervention and these outcomes.

The model shows direct links from the intervention to both teacher outcomes; these links capture the knowledge and practices teachers gain from exposure to the program. It also shows direct links from each of the teacher outcomes to student achievement. These links indicate that as teachers develop and improve their knowledge and practice of reading and instructional strategies, student learning and student achievement in reading improve.[21] The model also shows links between teacher knowledge and teacher practice, indicating that gains in the two areas reinforce each other.

[20] In a spiral learning approach, learners are exposed to concepts repeatedly through a progressive learning and relearning process. Concepts are first introduced at a basic level. Over time, they are revisited repeatedly with increasing sophistication to deepen learners' understanding.

[21] Interactions between program staff and students during the intervention—through both lesson demonstrations and instructional assistance in the classroom—may also affect student outcomes.

Figure 1.2 Theoretical model of Pacific CHILD



Source: Authors' construction.

Research questions and study design

Although the relationships in the model are based on the theoretical impact of Pacific CHILD on individual teachers and students, this study examines the effects of the intervention as a school-level program rather than an individual-level program. It investigates the effects of the intervention at the school level because Pacific CHILD is designed and implemented as a school-level intervention that emphasizes school-based group activities as well as in-classroom activities during regular school hours. These activities require the consent, if not active support, of school administrators. The study was designed to assess whether the intervention, as implemented in the field, had impacts on students (primary outcome) and teachers (secondary outcome) in schools that were offered the opportunity to participate in the intervention.

The two-year intervention provided professional development to teachers of grade 4 and grade 5 English language arts. Therefore, the study focuses on outcomes of grade 5 students and grade 4 and grade 5 teachers, who could have been potentially exposed to the full intervention at the end of two years.

This study investigated two confirmatory research questions:

- Did grade 5 students at schools that were offered Pacific CHILD for two years perform differently on assessments of reading comprehension from grade 5 students at schools that were not offered Pacific CHILD?
- Did grade 4 and grade 5 teachers at schools that were offered Pacific CHILD for two years perform differently from teachers at schools that were not offered Pacific CHILD, as measured by either an assessment of their knowledge of theories and strategies related to effective reading instruction (including English language learner-focused theories and strategies) or an assessment of their instructional practices for enhancing student reading comprehension (including English language learner-focused practices)?

These research questions reflect the focus on teachers and students at schools that were offered the intervention rather than on individuals who were offered the intervention.

A multisite, cluster random assignment design was adopted, in which schools were the unit of random assignment and teachers and students within schools were the units of analysis. Student and teacher domains were investigated independently. The primary purpose was to assess the effectiveness of the intervention in improving students' reading comprehension. The study would conclude that the intervention had an effect on students if the impact analyses demonstrated a statistically significant treatment effect on student achievement in reading comprehension at the end of the two-year intervention. The study would conclude that the intervention had an effect on teachers if the impact analyses demonstrated a statistically significant treatment effect on either teacher knowledge or teacher practice at the end of the two-year intervention.

As the intervention was designed to be effective across the diverse Pacific region, the focus of the study was on the overall program effects across the three study entities. Data from all entities were thus pooled in the confirmatory analysis. As part of the exploratory analyses, the study examined impacts by entity.

Content and organization of this report

Chapter 2 presents the study design in detail. It describes the recruitment and randomization process, baseline equivalence, the study sample, and the analysis framework. Chapter 3 summarizes data collection, describing the data collection instruments, procedures, and completion rates. Chapter 4 provides additional information on the structure of the professional development program and describes program implementation. Chapter 5 reports the findings from the impact analyses for student achievement, teacher knowledge, and teacher practice. Chapter 6 presents the results of exploratory analyses. Chapter 7 summarizes the study's findings and discusses its limitations. Appendixes provide details on the implementation of the Pacific CHILD program, the study sample, the study design, data collection, and statistical approaches.

Chapter 2: Study design

This chapter describes the random assignment design, recruitment process, sample definitions and descriptions, analytic framework, and methodology for the impact study.

Random assignment design

In randomized controlled trial studies, participants are randomly assigned to either a treatment group, which receives an intervention, or a control group, which does not. The impact of the intervention is measured as the difference in outcomes between the treatment and control groups. Randomized controlled trial studies allow for the causal inference of program impacts, because the process of random assignment creates equivalent groups in expectation that systematically differ only in their opportunity to receive the intervention.

This study was designed as a randomized controlled trial with the multilevel structure of a typical cluster random assignment design. The school was the unit of random assignment, reflecting the school-level nature of Pacific CHILD. The treatment condition consisted of offering schools the opportunity to participate in the two-year Pacific CHILD professional development program for teachers of grade 4 and grade 5 English language arts.[22] Before the intervention began, teachers at treatment group schools were actively recruited to join the program. Not all grade 4 and grade 5 teachers who joined program schools after the invention began were actively recruited into the program.[23] However, teachers in treatment group schools in any grade were free to participate in Pacific CHILD peer group meetings (known as *structured learning teams*) regardless of whether they formally participated in the intervention. As discussed below, not all teachers who were invited to participate in the intervention chose to take part in it.

Outcomes for students and teachers from the study schools were examined toward the end of the two-year intervention period. The study samples included all grade 5 students and grade 4 and grade 5 teachers who were at the study schools near the end of the second year of the intervention, regardless of whether or how much they had been potentially or actually exposed to the intervention. Thus, the impact study measured the intent-to-treat effects of Pacific CHILD as implemented in the field.

The control condition consisted of "business as usual". Schools in the control group did not participate in the Pacific CHILD program during the two years in which the intervention was implemented at treatment group schools (they were offered delayed treatment two years after the study ended). Except for an embargo on Pacific CHILD, the study did not impose any requirements or make any requests regarding professional development activities for teachers at control group schools: they were not restricted from using other services, such as support

[22] Teachers who were using a curriculum that was not compatible with Pacific CHILD were not targeted by the intervention and were considered ineligible for the study. For more information on which teachers were considered ineligible, see the section on student and teacher samples for the impact analysis.

[23] After the initial recruitment period, how actively new teachers were recruited depended on when they joined the schools, the capacity of program staff to serve additional participants, and other resource restrictions of the intervention.

services provided by curriculum developers, professional development provided by districts, or any other resources to which they might have had access on their own.

Given the clustered nature of the data, the study applied statistical modeling to estimate the standard errors for the estimates of program impact. Multivariate analyses were used to improve the precision of the impact estimates and to offset the effect of clustering on the study's statistical power. The a priori calculations of the sample size requirements for this study took into account the effects of clustering and made the assumption, based on findings from previous studies, that statistical power could be improved through statistical adjustments.[24] These a priori calculations led to a recruitment goal of 50 schools to attain the minimum detectable effect size of 0.15 for students and 0.40 for teachers.[25] (For additional information on statistical power and sample size calculations for the study, see appendix A.)

School recruitment and selection

This study is based on a convenience sample of public elementary schools in three entities in the Pacific: American Samoa, the Commonwealth of the Northern Mariana Islands (CNMI), and Hawai'i. The three entities were selected based on the availability of student outcome data and their willingness to participate in the study. Of the 10 entities in the Pacific Region served by REL Pacific, four administered comparable national, norm-referenced tests for the elementary grades at the time of site selection.[26] Three of the four entities, American Samoa, the CNMI, and Hawai'i, agreed to participate in the study.

Throughout the study, Berkeley Policy Associates (BPA), a third-party evaluator, worked independently of REL Pacific, the organization that developed and implemented the Pacific CHILD program. Recruitment was the one activity that required the collaboration of both teams. REL Pacific staff led the recruitment efforts; the BPA research team provided potential participants with technical information on the research design and the requirements for data collection activities. The BPA research team also worked with REL Pacific staff to develop guidelines for recruiting the target number of schools and the criteria for prioritizing which schools to target.

At the beginning of the recruitment process, the goal was to recruit a minimum of 50 schools, based on the a priori power analysis. In determining the target number of schools across entities and within each entity, the study took account of a number of factors, including the power needed to allow reliable student- and teacher-level estimation of program effects and to conduct subgroup analyses for Hawai'i as well as for the pooled sample; support from the local education

[24] Previous studies show that the effects of clustering on statistical power in random assignment studies can be reduced by controlling for cluster-level variation in estimating program impacts (see Schochet 2005; Bloom, Richburg-Hayes, and Black 2007; and Bloom, Bos, and Lee 1999).

[25] For the primary student outcome, the goal was to attain a minimum detectable effect size of at least 0.20, which is in line with other recent studies funded by the Department of Education. Because of resource constraints and the expectation that impacts on teachers, if any, would be larger than impacts on students (given that the intervention targets teachers directly), the study did not aim to attain the same level of power for the teacher outcomes.

[26] The 10 Pacific entities in the region are American Samoa; the CNMI; Guam; Hawai'i; the Federated States of Micronesia (Chuuk, Kosrae, Pohnpei, and Yap); the Republic of the Marshall Islands; and the Republic of Palau. Only American Samoa, the CNMI, Hawai'i, and Guam administered national, norm-referenced tests to grade 4 and grade 5 students at the time the study was designed.

agency staff; interest in the intervention by teachers; and the study's resource constraints.[27] The recruitment goal was at least 50 schools, including at least 26 schools in Hawai'i.[28]

In American Samoa, which has a centralized public education system, REL Pacific staff concentrated their recruitment efforts on obtaining support from the director and senior administrators at the American Samoa Department of Education. Together with BPA researchers, they worked with department administrators to guide recruitment decisions regarding issues such as the target number of schools and the minimum school size. The American Samoa Department of Education then selected schools for the study from the 23 public elementary schools in the entity and encouraged principals and teachers to participate in the study.

In the CNMI, which also has a centralized public education system, REL Pacific staff first gained the support of the commissioner of the CNMI Public School System, who encouraged all 12 public elementary schools in the entity to consider participating in Pacific CHILD. At schools that agreed to participate in the study, principals encouraged teachers to participate in the research and the intervention.

In Hawai'i, REL Pacific staff first sought support from state education officials. They then solicited help from the leadership of "complex areas" (local administrative areas within the Hawai'i Department of Education). Schools in complex areas with leadership support were ranked for preference based on their status under the No Child Left Behind Act of 2001, the proportion of English language learner students, and the proportion of students eligible for free or reduced-price lunch. These systematic rankings gave recruitment priority to schools that were not meeting the state's annual measurable objectives and that had high proportions of English language learner students and students eligible for free or reduced-price lunch.[29] REL Pacific staff recruited schools starting from the top of the list until the target sample size was reached, considering the geographical balance of the sample across islands and the level of interest in the intervention by teachers. To recruit schools in Hawai'i, REL Pacific staff and BPA researchers made presentations and answered questions. They followed up these visits with phone and email communications with administrators and teachers.

---

[27] The recruitment criteria for schools included interest in participating in the intervention from at least two teachers at each school. Expressions of interest among teachers at the time of recruitment did not necessarily translate into a commitment to participate in the study, as recruitment took place in the spring before the final determination of teacher assignments during the first study school year. School-level commitment to participate in the study was secured before random assignment with a formal memorandum of understanding signed by school administrators.

[28] The goal for Hawai'i was initially set at 25. It was revised to 26 because of the loss of schools recruited earlier. Appendix A provides information regarding this revision.

[29] Each school received scores in four categories: (a) the proportion of English language learner students (for all grades in the school); (b) the proportion of students eligible for free or reduced-price lunch (for all grades in the school); (c) current year annual yearly progress determination (whether the school met or did not meet its annual yearly progress); and (d) status under the No Child Left Behind Act of 2001: in good standing (met annual yearly progress for two or more years); corrective action (did not meet annual yearly progress for one to two years); school improvement (did not meet annual yearly progress for three to four years); planning for restructuring (did not meet annual yearly progress for five years); and restructuring (did not meet annual yearly progress for six or more years). All categories were scored using the most recent data available. The scoring system assigned a score ranging from zero to four points in each category, with higher scores given to schools with larger proportions of English language learner students, larger proportions of students eligible for free or reduced-price lunch, and less success in meeting annual yearly progress. Schools were ranked for recruitment based on their total score across the four categories. If the scores were tied, schools were ranked in the order of the score for the proportion of English language learner students, followed by annual yearly progress status and the proportion of students eligible for free or reduced-price lunch.

Fifty-one schools agreed to participate in the study: 26 in Hawai‘i and 25 in American Samoa and the CNMI (IES guidelines prevent the separate disclosure of the number of participating schools in each of these entities). In both American Samoa and the CNMI, the recruited schools were spread across the most populated island and an outer island, with the majority of the recruited study schools located on the most populated island. In Hawai‘i the recruited study schools were spread across the state's four major islands. In all three entities, support from school administrators was formalized in a memorandum of understanding that specified their commitment to permitting REL Pacific staff to lead program activities at their school, allowing teachers to participate in the program activities, and allowing researchers to conduct data collection activities. After schools were recruited, agreed to join the study, and completed the memorandum of understanding, teachers at participating schools still had the option to participate or not participate.

Recruitment activities took place during the 2006/07 and 2007/08 school years. In American Samoa and the CNMI, recruitment was completed during the 2006/07 school year, and the intervention started in the summer of 2007. Recruitment efforts in Hawai‘i during the 2006/07 school year did not yield the target number of schools. Outreach efforts in Hawai‘i therefore continued during the 2007/08 school year, and the intervention started in the summer of 2008.

Characteristics of recruited schools

On average, the 51 recruited schools served 139 grade 4 and grade 5 students and had a student-to-teacher ratio of 17 students per teacher (table 2.1). For the sample as a whole, 88.5 percent of students at these schools were of races/ethnicities other than White, and 72.5 percent were eligible for free or reduced-price lunch.

Table 2.1 Characteristics of recruited schools

Average number of students per school, all grades	538
Average student-teacher ratio, all grades	17:1
Average number of students in grades 4 and 5 per school	139
Average student-teacher ratio in grades 4 and 5	21:1
Percentage of students eligible for free or reduced-price lunch, all grades	72.5
Percentage of students of races/ethnicities other than White, all grades	88.5

Source: Authors' analysis based on data from U.S. Department of Education (2010a) for American Samoa and the Commonwealth of the Northern Mariana Islands and U.S. Department of Education (2010b) for Hawai‘i. Figures for students and teachers per school in grades 4 and 5 are based on enrollment estimates from the American Samoa Department of Education, the Commonwealth of the Northern Mariana Islands Public School System, and the Hawai‘i Department of Education.

Random assignment of schools

The recruited sample of 51 schools was randomly assigned to either a treatment or control condition. Schools were assigned using restricted randomization, in which recruited schools were blocked to ensure that the resulting allocation of schools in the treatment and control groups was balanced in terms of both the number of schools and key school characteristics.

In American Samoa, recruited schools were blocked based on location in a semi-urban area (in the vicinity of the capital city) and school size. American Samoa has a semi-urban center around the capital, which contrasts with the less densely populated villages that characterize the rest of the entity. School enrollment ranged from less than 100 to more than 1,000 students. Given the relatively small number of schools recruited, schools were blocked by location and size, in order to ensure that the geographical distribution and number of students and teachers was balanced across conditions.

In the CNMI, recruited schools were blocked by school size, which ranged from less than 300 to more than 600 students, in order to ensure that the number of students and teachers was balanced across conditions.

In Hawai‘i, recruited schools were blocked by island. Within each island, if multiple schools were recruited from one “complex area” (a local administrative area of the Hawai‘i Department of Education), they were grouped together to ensure that at least one school was assigned to the treatment condition, in order to maintain support from the complex area for this study. Random assignment was conducted in two phases, in order to notify schools of their assigned condition as soon as possible after recruitment in order to retain them in the study.[30]

Within each block, random assignment was carried out with an allocation ratio of 1:1 to ensure that randomization yielded the same number of schools in each condition: half the schools in each block were randomly assigned to the treatment group, and the other half were randomly assigned to the control group. This random assignment procedure resulted in a treatment group of 26 schools (13 schools in American Samoa and the CNMI and 13 in Hawai‘i) and a control group of 25 schools (12 in American Samoa and the CNMI and 13 in Hawai‘i).

After the randomization of schools, REL Pacific staff invited grade 4 and grade 5 teachers who taught English language arts at the treatment group schools to participate in the Pacific CHILD professional development program. Over the subsequent two years (the 2007/08 and 2008/09 school years in American Samoa and the CNMI and the 2008/09 and 2009/10 school years in Hawai‘i), REL Pacific staff delivered the professional development program to the teachers at the treatment group schools who volunteered to participate. The first year of the intervention is referred to as year 1; the second year is referred to as year 2. The school year immediately before the intervention is referred to as the baseline year. The research team collected data corresponding to the baseline as well as the first and second years of intervention from all study schools that were randomized.[31]

---

[30] The first group of 20 schools was randomly assigned in January 2008; the second group of six schools was randomly assigned in April 2008, before the first program activity took place in June 2008.

[31] Attempts were made to collect data from all study-eligible teachers from all schools originally randomly assigned, regardless of whether they actually participated in the intervention or research activities.

Adjustment to original sample

During the first year of implementation, the study team received a report that several treatment group teachers who were unable to attend the first annual institute were reassigned by school administrators to other grades within the same school. They were replaced with teachers from other grades within the same school. This change was brought to the attention of the study team after REL Pacific held the first annual institute.

While teacher reassignment after random assignment may be considered as a normal part of implementing a school-level intervention like Pacific CHILD, the possibility that these particular reassignment decisions were systematically carried out in response to the study treatment itself could not be ruled out.[32] Because such reassignment could potentially compromise the integrity of the experimental design, it was decided to exclude the affected schools from the impact analysis, together with other study schools in the same assignment blocks.[33] Their exclusion resulted in a reduction of the impact analysis sample by six schools (three treatment group schools and three control group schools) from the original 51 schools. When the study's statistical power was recalculated based on the reduced sample size assumption, the minimum detectable effect size for the 45 remaining schools was estimated to be 0.16 for students and 0.46 for teachers, compared with 0.15 for students and 0.40 for teachers based on the original sample size assumption. Based on the recalculation, it was concluded that the study with the reduced sample still maintained sufficient power to produce meaningful findings (for additional information on the power analysis for the reduced sample, see appendix A). The remaining 45 schools (23 treatment schools and 22 control schools) constitute the impact sample referenced in the rest of this report.[34]

Characteristics of schools in impact sample

The characteristics of the 45 schools in the impact sample are presented in table 2.2, along with the characteristics of the 51 schools in the original sample. The six schools excluded from the study were larger than average schools. As a result, the average number of students in the schools retained for impact analyses decreased once these schools were excluded.

---

[32] For example, under usual non-study environments, schools may not have replaced any eligible teachers even if there were schedule conflicts with an intervention; or may have still reassigned teachers due to teachers' schedule conflicts but using different criteria for finding a replacement teacher.

[33] While dropping the control schools in the same blocks may not completely address potential bias due to dropping the affected program schools or may potentially introduce other types of bias, the study team chose to remove the entire blocks to balance the sample across the conditions with respect to known observable criteria used to block the schools.

[34] Although the affected schools were excluded from the confirmatory impact analysis, no changes were made to the implementation of Pacific CHILD or to data collection for the evaluation.

Table 2.2 Characteristics of schools in impact sample, by entity

Characteristic	Original sample	Impact sample			
	Total	Total	American Samoa	Commonwealth of the Northern Mariana Islands	Hawai‘i
Average number of students per school, all grades	538	475	200	484	566
Average student-teacher ratio, all grades	17:1	16:1	14:1	20:1	15:1
Average number of students in grades 4 and 5 per school	139	129	42	141	155
Average student-teacher ratio in grades 4 and 5	21:1	20:1	16:1	20:1	22:1
Percentage of students eligible for free or reduced-price lunch, all grades	72.5	69.0	98.1	99.3	47.3
Percentage of students of races/ethnicities other than White, all grades	88.5	86.9	100.0	99.6	77.6
Number of schools	51	45	19[a]		26

a. Following Institute of Education Sciences guidelines, the numbers of schools in American Samoa and the CNMI are combined to prevent disclosure risk.

Source: Authors' analysis based on data from U.S. Department of Education (2010a) for American Samoa and the Commonwealth of the Northern Mariana Islands and U.S. Department of Education (2010b) for Hawai‘i. Figures for students and teachers per school in grades 4 and 5 are based on enrollment estimates from the American Samoa Department of Education, the Commonwealth of the Northern Mariana Islands Public School System, and the Hawai‘i Department of Education.

On average, schools in the impact sample served 129 grade 4 and grade 5 students and had 16 students per teacher during the baseline year. On average, 86.9 of the students at these schools were of races/ethnicities other than White, and 69.0 percent were eligible for free or reduced-price lunch.

The average profile of the schools in the impact sample varied across entities. Most notably, the average number of students at schools in American Samoa was less than half that in the CNMI or Hawai‘i: The average number of students per school was 200 in American Samoa, 484 in the CNMI, and 566 in Hawai‘i. The differences in the school size were partly a result of the removal of six schools from the impact study, which included larger schools in American Samoa. Cluster (school) size differences could potentially lead to variation in how precisely the impacts could be estimated across the study entities. As discussed later in this chapter and in chapter 5, this concern was taken into account in refining the estimation approach.

In the impact sample schools in American Samoa and the CNMI, on average more than 90 percent of students were eligible for free or reduced-price lunch (table 2.2). In the sample schools in Hawai‘i, an average of 47 percent of students were eligible for free or reduced-price lunch. The average percentage of White students in sample schools was less than 1 percent in American Samoa and the CNMI and about 22 percent of students in Hawai‘i. The average

student-teacher ratio based on all grades in the sample schools was 16 in American Samoa, 20 in the CNMI, and 22 in Hawai‘i.

Average school characteristics at baseline thus differed across entities. The outcome measures could be correlated with these school characteristics and other baseline factors that varied by entity. The study accounted for such factors in the impact estimation. Even with variation across entities, however, the study was expected to yield unbiased estimates of impacts, because schools were randomly assigned to either the treatment or control group within entities.

Baseline comparison of treatment and control groups

To evaluate whether random assignment resulted in statistically equivalent groups at baseline, the study compared selected school-level baseline characteristics of the treatment and control group schools in the impact sample (table 2.3). School-level data from the Common Core of Data, enrollment records, and student test records were collected to check baseline equivalence.[35] School-level characteristics included school size, the student-teacher ratio, the percentage of students eligible for free or reduced-price lunch, student race/ethnicity, and student achievement in reading. In Hawai‘i, the proportion of English language learner students was also compared (reliable official data on English language learner status were not available for American Samoa or the CNMI).[36] Student test records at baseline were available for grade 5 students in the CNMI and Hawai‘i and for grade 4 students in American Samoa and Hawai‘i. The two grades were compared separately.[37] Characteristics were compared at the school level by averaging the characteristics within schools, which served as the unit of random assignment. Treatment and control groups did not differ from each other in a statistically significant way on any of the baseline characteristics examined.

[35] The Common Core of Data database, compiled annually by the U.S. Department of Education’s National Center for Education Statistics, contains school-, district-, and state-level fiscal and nonfiscal data on public schools in the United States and its territories.

[36] The research team sought information on English language learners directly from school administrators. Because of the lack of a consistent definition of English language learners used in these entities, these data are not presented.

[37] The student assessment data were based on national, norm-referenced tests that were designed to be vertically scaled to allow comparisons of grade 4 and 5 scale scores.

Table 2.3 Baseline characteristics for treatment and control group schools

Baseline characteristic	Mean (standard deviation) Overall	Treatment schools	Control schools	Estimated difference	Test of difference *p*-value
Number of grade 4 students	65.0 (40.0)	64.7 (42.5)	65.3 (38.2)	–.6	.962
Number of grade 5 students	64.1 (41.4)	64.2 (42.5)	64.0 (38.2)	.3	.983
Number of grade 4 and grade 5 teachers	6.0 (3.1)	6.0 (3.4)	6.0 (2.9)	–.0	.962
Student-teacher ratio in grades 4 and 5	20.2 (4.8)	20.1 (4.9)	20.3 (4.9)	–.1	.935
Proportion of students eligible for free or reduced-price meals, all grades	69.0 (28.7)	69.8 (28.5)	68.1 (29.6)	1.7	.842
Mean proportion of students of races/ethnicities other than White, all grades	86.9 (15.2)	87.7 (13.9)	86.1 (16.8)	1.6	.722
Mean proportion of English language learner students (Hawai‘i only), all grades[a]	12.8 (9.7)	13.5 (8.2)	12.3 (11.4)	1.2	.764
Mean reading comprehension score (SAT 10 scale score),[b, c] grade 4	609.6 (20.5)	608.9 (21.1)	610.4 (20.6)	–1.5	.829
Mean reading comprehension score (SAT 10 scale score),[b, d] grade 5	636.5 (11.0)	634.4 (11.9)	638.6 (9.9)	–4.2	.260
Number of schools	45	23	22		
Number of schools in Hawai‘i	26	13	13		

Note: Significance tests are based on two-tailed *t*-tests, accounting for clustering at the school level.

a. Data on English language learner status were available only for Hawai‘i students.

b. TerraNova reading comprehension scores from Hawai‘i were converted to estimated Stanford 10 reading comprehension equivalents using published norming tables and concordancing method (see appendix E).

c. At baseline, grade 4 students in American Samoa and Hawai‘i completed standardized assessments. Thirty-five schools (18 treatment group schools and 17 control group schools) had grade 4 scores.

d. At baseline, grade 5 students in the CNMI and Hawai‘i completed standardized assessment. Thirty six schools (18 treatment group schools and 18 control group schools) had grade 5 scores.

Source: Authors' analysis based on data from U.S. Department of Education 2010a for American Samoa and the Commonwealth of the Northern Mariana Islands, U.S. Department of Education 2010b for Hawai‘i. Figures for students and teachers per school in grades 4 and 5 are based on enrollment estimates from the American Samoa Department of Education, the Commonwealth of the Northern Mariana Islands Public School System, and the Hawai‘i Department of Education.

Student and teacher samples for the impact analysis

This section describes the students and teachers investigated by this study. It first discusses how the student and teacher samples were defined for impact analyses. It then describes the samples and examines the turnover of students and teachers at the study schools. The section concludes by summarizing the profiles of the student and teacher samples.

Student and teacher impact samples

The student impact sample was defined as all grade 5 students enrolled in study schools at the time of testing in the spring of the second year of the intervention. The teacher impact sample was defined as all grade 4 and grade 5 teachers who taught English language arts in a self-contained classroom at the study schools in the spring of the second year of the intervention. Because of the intervention's focus on reading skills, only teachers who regularly taught English language arts were included in the teacher impact sample.[38] The sample excluded instructors who were not targeted by the intervention, such as teachers' aides, librarians, and teachers who taught only physical education, music, science, math, foreign languages, and other specialized subject areas. The criteria for teacher inclusion in the impact sample was the same as the criteria for recruiting teachers to the intervention (that is, grade 4 and grade 5 teachers who regularly taught English language arts), except that the impact sample was limited to teachers who were at the study schools in the spring of the second year of the intervention. The student and teacher samples used in the impact analyses were independently defined and analyzed. The student sample was not intended to represent the exact cohort taught by the teachers in the impact sample.

The student and teacher impact samples were defined so as to allow the assessment of Pacific CHILD as a two-year intervention. By the time the outcome data were collected, in the second year of the intervention, treatment group schools had been offered the intervention for nearly two full years; grade 5 students and grade 4 and grade 5 teachers at these schools could have been exposed to the intervention for up to two years. Analyses based on samples allow inferences to be made about the effectiveness of Pacific CHILD on students and teachers at schools where the intervention was implemented for nearly two years.

Cross-sectional nature of student and teacher samples

The study did not establish a fixed sample of students or teachers at the time of random assignment to be tracked through the study period. Instead, it defined the student sample as all grade 5 students who were enrolled at the study schools at the time of data collection (testing) during spring of the second year.[39] Similarly, the teacher sample was defined as all grade 4 and grade 5 teachers who were at the study schools at the time of data collection (in the spring of the second year) and met the criteria above. Thus, "cross-sectional" groups of students and teachers

[38] Not all regular classroom teachers at the study schools taught English language arts. In addition, a small number of teachers (fewer than five) in the impact sample were required by their school to implement completely scripted English language arts lessons and could not use reading and instructional strategies emphasized in Pacific CHILD. Those teachers were excluded from the study before random assignment.

[39] Consistent with the study's intention to estimate intent-to-treat impacts of this school-level intervention, all grade 5 students in the study schools were included, even students whose classroom teachers did not participate in the intervention.

who were selected at one point in time after random assignment constituted the impact samples of this study.[40]

The study did not directly estimate the effects of Pacific CHILD on individuals who were exposed to the intervention for nearly two years. Rather, it aimed to investigate the intent-to-treat impacts on teachers and students who could have been exposed to Pacific CHILD while their school participated in the intervention. The impacts estimated can be interpreted as measuring whether Pacific CHILD would make a difference in student and teacher outcomes at schools where the intervention had been implemented for two years. This interpretation is consistent with the study's focus on the school-level nature of the intervention.

Student and teacher movement in and out of the study schools

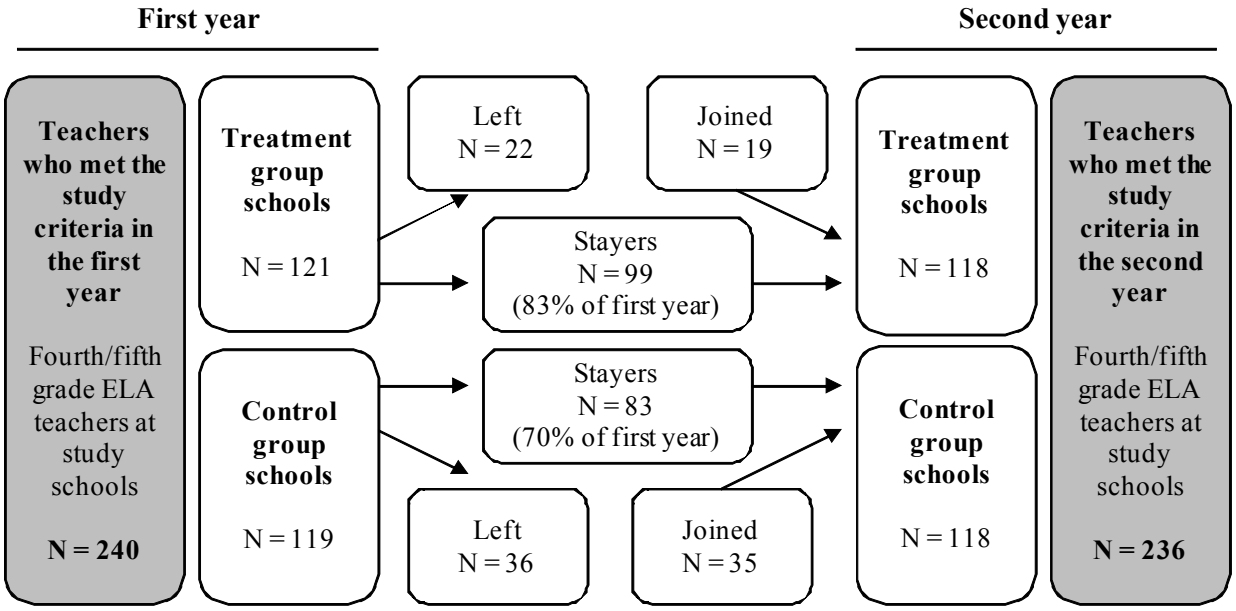
The student and teacher impact samples were cross-sectional samples determined near the end of the two-year intervention. As a result, the usual concerns about attrition over the course of the study period do not directly apply. The movement of students and teachers at the study schools over time is nevertheless reviewed here, in order to assess potential differential turnover across conditions. Any movement of individuals—especially teachers, the primary target of the intervention—could have been partially induced by the intervention.

During the intervention period, the mobility of teachers in and out of the study schools was documented. The majority of teachers in the study schools who met the study criteria in the first year met the study criteria and remained at the study schools in the second year (figure 2.1). No teachers in the impact sample transferred across conditions. In schools in the treatment group, 99 of 121 (82 percent) of teachers remained grade 4 or grade 5 teachers across both years; 83 of 119 (70 percent) of the teachers in the control group schools did so. Overall, turnover of teachers was lower in treatment group schools than in control group schools. This difference is statistically significant at the 5 percent level ($p = .029$). Further inspection of teacher turnover by entity showed that the difference across conditions was driven by the movement of teachers in one entity, in which at least one control group school with an unusual circumstance unrelated to the intervention contributed to the higher turnover.[41]

[40] Information on all grade 4 and grade 5 teachers who had been at the study schools during the intervention period was collected, regardless of whether they were part of the impact sample. These data were collected in order to monitor movement between treatment and control group schools, document participation patterns, and measure the dosage of intervention received by treatment group teachers in the impact sample.

[41] The difference in turnover rates (defined as the ratio of leavers to the first year total) was statistically significant at the 5 percent level in only one of the three entities ($p = 0.025$). Consistent results were found based on alternative definitions of turnover rates, including the ratio of leavers to the average of the first year and second year totals and the ratio of stayers to the second year total.

Figure 2.1 Movement of teachers at study schools



Source: Authors' construction based on data collected for study.

The turnover of students at the study schools was estimated using the state standards test, as available. The data used to estimate student turnover included only students who completed state tests; they did not include all students enrolled in the study schools. Differential turnover patterns between the treatment and control group students were not detected (table 2.4). The average estimated year-to-year turnover rate was about 16 percent for the sample; an estimated 75 percent of the grade 5 students in the target sample had been continuously enrolled in the same study schools from baseline to the end of the second year of the intervention (76 percent for treatment group schools, 75 percent for control group schools, $p = .310$). During the same period, an estimated 3 percent of students crossed conditions, with no statistically significant differences by study condition ($p = .588$).

The composition of the impact analysis samples was affected by movement of individuals in and out of the study schools during the intervention period. Differential movement across conditions could lead to comparison groups that were no longer equivalent in expectation, raising concerns about potential bias in impact estimates. The goal of this study was to assess the impacts of the school-level intervention as implemented in the field, however. Any intervention-induced movement of individuals could thus be considered part of the treatment condition. Readers are cautioned to consider potential bias and implications of turnover in evaluating the teacher impact analysis results.

Table 2.4 Movement of students at study schools

Movement indicator	All schools	Treatment schools	Control schools	Test of difference p-value
Enrolled in same school for two years (percent of baseline year cohort)	75.4	76.2	74.6	.310
Annual turnover at study schools[a]	15.7	14.9	16.4	.066
Crossed over conditions (percent of baseline year cohort)	3.3	3.4	3.1	.588
Average number of students (per year)	3,088	1,580	1,508	na

na is not applicable.

a. Turnover per year was calculated as the ratio of the number of students who left between the baseline and the second year to the average of the number of students at baseline and the second year, divided by two to annualize.

Source: Authors' analysis based on state test records provided by the American Samoa Department of Education, the Commonwealth of the Northern Mariana Islands Public School System, and the Hawai'i Department of Education.

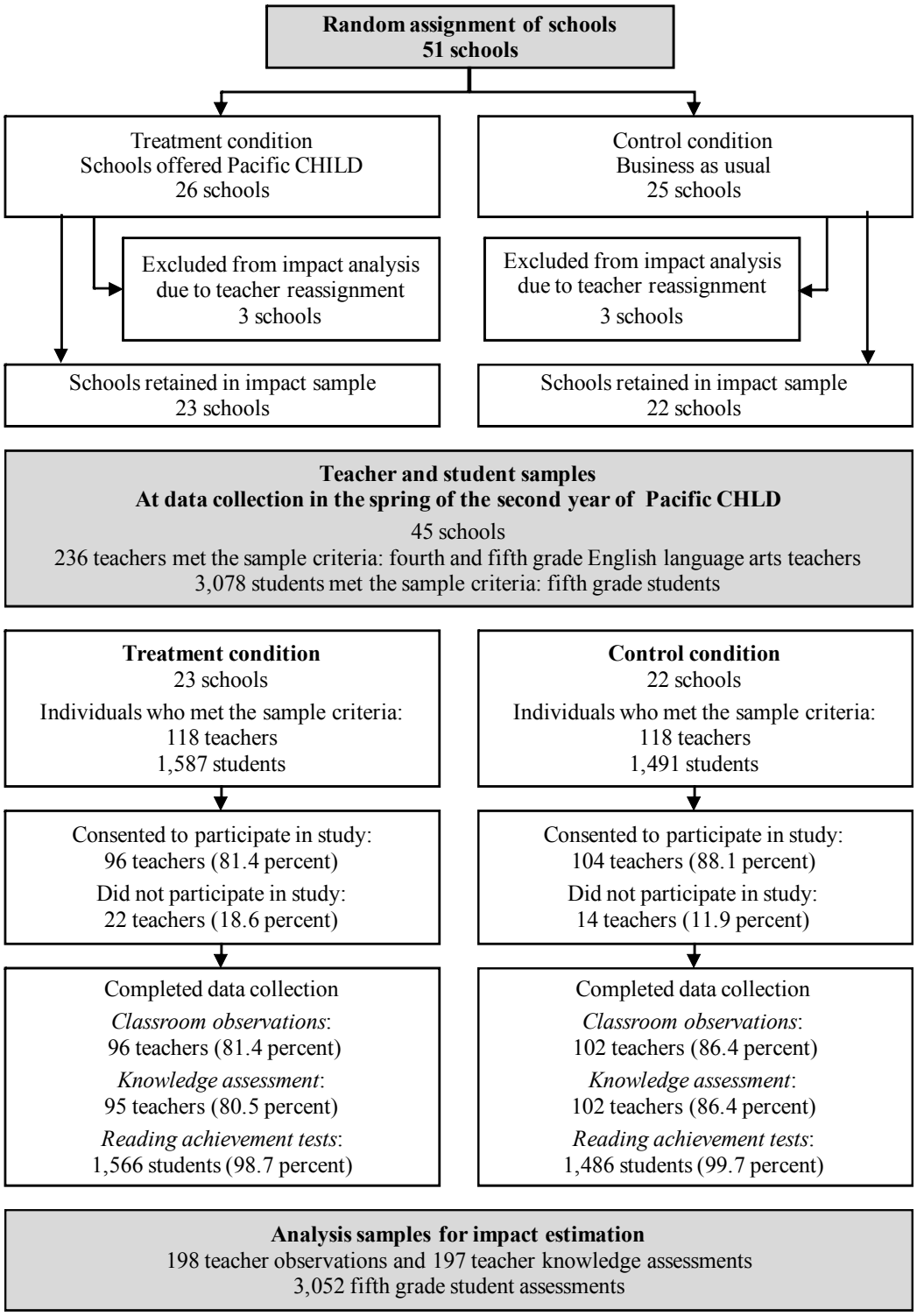
Description of student and teacher analysis samples

In total, 3,078 grade 5 students and 236 grade 4 and grade 5 teachers met the sample criteria. Students and teachers who met the sample criteria for whom outcome data were collected make up the analysis samples. The analysis samples included 3,052 students for the assessment of impacts on achievement in reading comprehension, 197 teachers for the assessment of the impacts on teacher knowledge, and 198 teachers for the assessment of impacts on teacher practice.

The analysis samples represent all targeted students at study schools except students who did not complete the assessments at the end of year 2 (nonrespondents) and all targeted teachers at study schools except teachers who did not consent to participate in any part of the study (nonconsenters) and teachers who consented to participate in the study but did not complete certain data collection activities (nonrespondents).[42] Figure 2.2 shows the extent of nonconsent and nonresponse.

[42] The state education agencies provided the research team with the student-level records, waiving parental consent. The data were provided under terms that meet the Family Educational Rights and Privacy Act (FERPA) as well as the entity's own data security requirements. The research team collected the teacher data directly from teachers.

Figure 2.2 Flow chart of sample selection, from random assignment of schools to selection of impact analysis sample



Source: Authors' construction, adapted from the Consolidated Standards of Reporting Trials (CONSORT) diagram suggested by Campbell, Elbourne, and Altman (2004), based on data collected for this study.

The recruitment of teachers at treatment group schools into the intervention and the recruitment of teachers into the study were separate activities. Teachers were assured that their consent (or refusal) to participate in the study would not mean consent (or refusal) to participate in the intervention and vice versa. In theory, it was possible for teachers to have agreed to participate in the intervention but then declined to participate in the study. In practice, no teachers who chose to participate in the intervention declined to participate in the study.

The extent of data completion among teachers reflects both their consent to participate in the collection of their outcome data as well as their actual response to the outcome data collection efforts. The difference in data completion rates between teachers in the treatment and control groups primarily reflected differences in the rates of consent to participate in the study. This difference, for example, could be due to their knowledge of whether or not the school was offered the intervention at the time when they were invited to participate in the study. As shown in figure 2.2, 22 of 118 of teachers at treatment group schools (18.6 percent) and 14 of 118 teachers at control group schools (11.9 percent) did not consent to participate in the study. In other words, the consent rate among the impact sample teachers was 81.4 percent in treatment schools and 88.1 percent in control schools. The difference was not statistically significant (p = .148). Because no data were collected from teachers who did not consent to participate in the study, reasons for nonparticipation in the study could not be systematically investigated. Chapter 3 presents additional information on the data completion and response rates for each outcome measure.

Sample equivalence and sample attrition

Systematic differences in the individual impact analysis samples across the conditions could lead to potential bias in the impact estimates. As illustrated in figure 2.2 and discussed above, this was a concern for teachers in this study. Given that the impact teacher sample was a cohort defined toward the end of the two-year intervention, factors such as teachers' knowledge of their schools' assignment status and their own exposure to the intervention to date—i.e., factors that differed across the conditions—could have influenced their participation in the outcome data collection efforts. If treatment group teachers who selected to participate in the outcome data collection efforts differed both in observed and unobserved ways from control group teachers who selected to participate, the assumption of sample equivalence in expectation across the conditions could be compromised, leading to potential bias in the study results.

One indicator for determining whether the equivalence between control and treatment conditions was compromised is sample attrition. Minimizing the overall attrition level and systematic differences in attrition rates across the conditions is critical for attaining reliable impact estimates because it increases the analysis sample size and decreases the risk of bias caused by missing data. Attrition in this study can be discussed at school and individual levels (see figure 2.2). At the school level, attrition is measured by the reduction in the sample over time. No school dropped out of the study throughout the study period. Thus, there was no attrition at the school level. At the individual (teacher or student) level, attrition was assessed in terms of the rates of data collection completion.

For the student sample, missing outcome data were minimal and balanced across conditions. Data completion rates for the 3,078 students in the sample were estimated to be 99.2 percent (98.7 percent for the treatment group and 99.7 percent for the control group) (see chapter 3).

Estimated overall attrition rate was thus 0.8 percent and the differential attrition rate 1.0 percent. The *What Works Clearinghouse Procedures and Standards Handbook* (U.S. Department of Education 2008) adopts an effect size of 0.05 standard deviation or less of the outcome measure as an acceptable level of attrition bias in impact estimates in applying its attrition bias model. According to the attrition bias model described in the What Works Clearinghouse handbook, the combination of the overall and differential attrition rates for the student data (a 0.8 percent overall rate and a 1.0 percent differential rate) resulted in a level of bias that is well below the attrition bias threshold.

For the teacher sample, of the 236 study-eligible teachers, 83.5 percent (80.5 percent in the treatment group and 86.4 percent in the control group) completed the teacher knowledge assessment, and 83.9 percent (81.4 percent in the treatment group and 86.4 percent in the control group) responded to the teacher practice observations (see figure 2.2 and chapter 3). For the knowledge assessment, the overall attrition rate was 16.5 percent, and the differential attrition rate was 5.9 percent. For the classroom practice observations, the overall attrition was 16.1 percent, and the differential attrition was 5.0 percent. Based on the attrition bias model in the What Works Clearinghouse handbook, the combination of the overall and differential attrition rates for the teacher practice data resulted in the acceptable level of bias. However, the combination of the overall and differential attrition rates for the teacher knowledge data resulted in a level of bias that is considered potentially acceptable.[43] The study attempted to minimize potential bias by applying statistical adjustments to impact estimation. Given the potentially acceptable level of bias for the teacher knowledge assessment, readers are cautioned to consider possible attrition bias in teacher impact analysis results.

Profiles of the student and teacher samples

The student sample profile was based on individual-level demographic information provided by school districts, along with reading comprehension scores on the assessment administered in the spring of the second year of the study. Data for the student profiles, along with reading achievement data, were available for 3,052 grade 5 students in the impact sample.

The teacher sample profile was based on the teacher background survey completed by 196 teachers in the impact sample during the spring of the second year of the study. These data, collected after random assignment, were used to describe the teacher characteristics that were time invariant and independent of the intervention (for example, gender, race/ethnicity) as well as characteristics that were considered to be stable over time and could be regarded as independent of the intervention within the span of the study period (for example, years of experience, level of education).[44]

[43] The acceptable bias threshold of 0.05 in effect size was applied to both student and teacher data. The threshold of 0.05 may be considered conservative for the teacher outcome data, for which larger effects were expected than for students.

[44] It is possible that some of these characteristics could be influenced by the assignment condition (for example, whether or not the teacher completed a degree program).

Table 2.5 Profile of student analysis sample, by entity

(percent, except where otherwise indicated)

Characteristic	All entities	American Samoa	Commonwealth of the Northern Mariana Islands	Hawai‘i
Female	47.7	42.7	47.8	48.1
Race/ethnicity other than White	88.2	100.0	99.7	83.5
Special education student	10.6	13.5	8.4	11.1
English language learner student	—	—	—	13.8
Eligible for free or reduced-price lunch	—	—	—	56.7
Number of students	3,052	185	692	2,175

— is not available.

Note: Percentage distributions are calculated based on completed responses.

Source: Authors' analysis based on student records provided by the American Samoa Department of Education, the Commonwealth of the Northern Mariana Islands Public School System, and the Hawai‘i Department of Education. Districts provided information on special education status in the student-level files.

The student sample was 47.7 percent female and 88.2 percent races/ethnicities other than White; 10.6 percent of the sample received special education services (table 2.5). For the Hawai‘i student sample, additional information was available on the percentage of students designated English language learner (13.8 percent) and eligible for free or reduced-priced lunch (56.7 percent).[45]

The teacher analysis sample surveyed was 80.6 percent female and 39.8 percent White (table 2.6). Across the three entities, 94.9 percent of teachers had at least a bachelor's degree, and 31.6 percent had an advanced degree. A large majority (87.7 percent) reported English as a primary language. On average, teachers had 8.7 years of experience teaching and had served 4.7 years at the current study school (at the baseline year).

[45] For American Samoa and the CNMI, reliable and comparable individual student-level data were not available in the records collected.

Table 2.6 Profile of teacher analysis sample, by entity

(percent, except where otherwise indicated)

Characteristic	All entities	American Samoa	Commonwealth of the Northern Mariana Islands	Hawai‘i
Female	80.6	61.1	79.0	84.5
Race, not mutually exclusive categories				
Pacific Islander	35.2	> 80.0	51.6	19.8
Asian	36.2		27.4	44.8
White, non-Hispanic	39.8	< 20.0	19.4	56.0
Primary language not English	12.3	66.7	13.1	3.4
Highest degree				
Associate’s degree	5.1	55.6	0.0	0.0
Bachelor’s degree	63.3	44.4	69.4	64.7
Advanced degree	31.6		30.6	35.3
Mean years of teaching at baseline	8.7	5.9	7.9	9.5
Mean years of teaching at study school at baseline	4.7	2.3	4.4	5.3
Number of teachers who completed survey	196	18	62	119

Note: Percentage distributions are calculated based on completed responses. Percentages for mutually exclusive race categories do not sum to 100. Other percentages may not sum to 100 because of rounding. Per Institute of Education Science guidelines, cells containing fewer than three units of observations are aggregated to protect confidentiality.

Source: Authors’ analysis based on teacher background survey conducted at end of second year of study.

Teachers in American Samoa reported lower levels of education than the sample as a whole: just 44.4 percent of teachers reported having at least a bachelor’s degree (all teachers in the CNMI and Hawai‘i reported having at least a bachelor’s degree). A larger percentage of teachers in American Samoa also reported speaking a language other than English as their primary language (66.7 percent in American Samoa, compared with 13.1 percent in the CNMI and 3.4 percent in Hawai‘i).

The percentage of teachers self-identifying as White was higher in Hawai‘i (56.0 percent) than in American Samoa and the CNMI (less than 20 percent). The average years of teaching experience was highest in Hawai‘i and lowest in American Samoa (9.5 years in Hawai‘i , 7.9 years in the CNMI, and 5.9 years in American Samoa).

Comparison of treatment and control groups

This section evaluates whether the student and teacher impact samples, defined at the time of data collection, represented statistically equivalent groups of individuals across conditions in terms of their background characteristics. To do so, the study examined a set of characteristics using the same data used to describe the student and teacher profiles in tables 2.5 and 2.6.

Although these data were collected after random assignment, they are regarded as invariant or stable over time and independent of the intervention. For this reason, they were used to approximate baseline characteristics. Together with baseline school-level characteristics, these characteristics were used as covariates in the impact estimation models to improve the precision of the impact estimates.

There were no statistically significant differences in the characteristics of the treatment and control groups in the student sample (table 2.7). For the teacher sample, the only statistically significant difference was in years of teaching (7.4 years for the treatment group and 9.9 years for the control group) (table 2.8). To address the difference across conditions, the study included years of teaching as a covariate in estimating the teacher impacts.

Table 2.7 Characteristics of students at treatment and control group schools

(percent, except where otherwise indicated)

	Mean (standard deviation)				
Characteristic	**Overall**	**Treatment schools**	**Control schools**	**Estimated difference**	**Test of difference *p*-value**
Female	47.7	46.7	48.8	–2.1	.244
	(.9)	(1.3)	(1.3)		
Race/ethnicity other than White	88.2	88.3	88.0	.3	.952
	(.6)	0.8)	(.8)		
Special education student	10.6	10.5	10.7	–0.2	.931
	(.6)	(.8)	(.8)		
English language learner student[a]	13.8	14.1	13.5	.6	.882
	(.5)	(.7)	(.8)		
Eligible for free or reduced-price lunch[a]	56.7	56.1	57.3	–1.2	.891
	(.9)	(1.2)	(1.3)		
Number of students	3,052	1,566	1,486		
Number of students in Hawai‘i	2,175	1,073	1,102		

Note: Table reports the standard deviation of a sample proportion estimated by the standard error of the proportion. Significance tests account for clustering at the school level. Significance tests are based on two-tailed t-tests for continuous variables and chi-square tests for categorical variables, accounting for clustering at the school level

a. Data available only for Hawai‘i students.

Source: Authors' calculations based on student records provided by the American Samoa Department of Education, the Commonwealth of the Northern Mariana Islands Public School System, and the Hawai‘i Department of Education.

Table 2.8 Characteristics of teachers at treatment and control group schools

(percent, except where otherwise indicated)

	Mean/proportion (standard deviation)				
Characteristic	**Overall**	**Treatment schools**	**Control schools**	**Estimated difference**	***p*-value**
Female	80.0 (2.8)	80.0 (4.1)	81.1 (3.9)	–1.2	.861
Race/ethnicity(non–mutually exclusive categories)					
Asian	36.2 (3.4)	35.8 (4.9)	36.6 (4.8)	–.8	.928
Pacific Islander	35.2 (3.4)	36.8 (4.9)	33.7 (4.7)	3.2	.730
White, non-Hispanic	39.8 (3.5)	40.0 (5.0)	39.6 (4.9)	.4	.972
Primary language not English	12.3 (2.4)	14.9 (3.7)	9.9 (3.0)	–5.0	.516
Highest degree					
Associate's degree	5.1 (1.6)	7.4 (2.7)	3.0 (1.7)	4.4	.574
Bachelor's degree	63.3 (3.4)	58.9 (5.0)	67.3 (4.7)	–8.4	
Advanced degree	31.6 (31.6)	33.7 (4.8)	29.7 (14.5)	4.0	
Years of teaching at baseline	8.7 (8.0)	7.4 (7.2)	9.9 (9.9)	–2.5*	.039
Years of teaching at study school at baseline	4.7 (5.7)	4.2 (5.6)	5.3 (5.3)	–1.1	.257
Taught grade 4 when outcome data were collected	52.6 (3.6)	53.7 (5.1)	51.5 (5.0)	2.2	.758
Number of teachers	196	95	101		

*Significantly different from zero at the .05 level (two-tailed test).

Note: Table reports the standard deviation of the sample proportion estimated by the standard error of the proportion. Significance tests are based on two-tailed *t*-tests for continuous variables and chi-square tests for categorical variables, accounting for clustering at the school level.

Source: Authors' analysis of teacher background survey conducted for this study.

Analysis framework

The basic framework of the study was to draw testable hypotheses from the research questions and to apply appropriate statistical testing to make inferences about the effectiveness of the intervention with respect to student achievement in reading comprehension (the primary research question) and teacher knowledge and practice (the secondary research question).[46] The following null hypothesis was tested to answer the primary confirmatory research question:

- Grade 5 students at schools assigned to Pacific CHILD did not perform differently from grade 5 students at schools not assigned to Pacific CHILD on standardized assessments of reading comprehension.

Rejection of this hypothesis would imply that Pacific CHILD had a statistically significant impact on student academic achievement in reading. A statistically significant impact would support the conclusion that Pacific CHILD affected student achievement.

To answer the secondary research question, the study tested the null hypothesis that teachers at schools assigned to the intervention did not performed differently from teachers at schools not assigned to the intervention on either the assessment of their knowledge or the assessment of their classroom practice. This hypothesis was investigated by testing the following domain-specific null hypotheses separately:

- Grade 4 and grade 5 teachers at schools assigned to Pacific CHILD did not perform differently from teachers at schools not assigned to Pacific CHILD on assessments of their knowledge of theories and strategies for reading comprehension and effective instruction.
- Grade 4 and grade 5 teachers at schools assigned to Pacific CHILD did not perform differently from teachers at schools not assigned to Pacific CHILD on assessments of their classroom instructional practice.

Separate statistical tests for the two teacher domains were conducted, adjusting *p*-values for multiple comparisons based on the method proposed by Benjamini and Hochberg (1995).[47] Rejection of either of the domain-specific hypotheses would imply that Pacific CHILD had a statistically significant impact on at least one of the two teacher outcomes. These hypotheses were derived from a theoretical model that predicts effects on individual teachers who were exposed to the intervention.

This study did not directly address the question of whether there was a difference between an individual teacher who participated in the intervention and a teacher who did not (or whether there was a difference between students of participating teachers and students of nonparticipating teachers). Instead, it examined whether there was a difference between teachers at schools offered the intervention and teachers at schools not offered the intervention (or whether students

[46] Although the study investigated impacts in student and teacher domains separately, the underlying theoretical model does not support specific conclusions about the overall program effectiveness based on impacts in any particular domain. Thus the effectiveness of Pacific CHILD is discussed with regard to the student outcome domain and the teacher outcome domains independently. Although some readers may be interested in drawing a single overall conclusion about program effectiveness based on impacts across domains, the study views the question of overall effectiveness as a policy inquiry, as it requires readers to define what overall effectiveness means based on their interests and priorities.

[47] See appendix B for a description of the Benjamini-Hochberg procedure.

enrolled at schools that participated in the intervention differed from students enrolled at schools that did not participate in the intervention).

The hypotheses tested were derived from a theoretical model that predicted effects on individuals who were exposed to the intervention. This study, however, did not directly address the question of whether there was a difference between individuals who were exposed to the intervention and those who were not. Instead, it examined whether there was a difference between individuals at schools offered the intervention and those at schools not offered the intervention.

Outcome measures

The impact on student achievement was measured using the reading comprehension subtests of national, norm-referenced tests administered in each entity as part of their regular assessment (the Stanford 10 Achievement Test [SAT 10] in American Samoa and the CNMI and the TerraNova, 2nd Edition, in Hawai‘i). For the impact analyses, TerraNova reading comprehension scale scores from Hawai‘i were converted to estimated SAT 10 reading comprehension equivalents using published norming tables and equipercentile methods. (For details on the linking of the two tests, see appendix E.)

Impacts on teacher knowledge were measured using a written teacher knowledge assessment developed for this study (see chapter 3). Teachers' instructional practices were measured through classroom observations using a modified version of the Sheltered Instruction Observation Protocol (SIOP®), expanded to include items relevant to Pacific CHILD (see appendix D).

Pooling of data across entities

Impacts were estimated by pooling data from the three entities, because the purpose of the study was to investigate whether Pacific CHILD, which was specifically designed to be adaptable to local contexts, yielded statistically significant impacts across the entities, rather than for a particular entity. As random assignment was conducted within entities, differences in average outcomes between treatment and control groups would still yield unbiased estimates of impacts overall, regardless of the extent of potential sample heterogeneity across entities.

The three entities were purposively selected as the study sites based on the availability of student outcome data and the administrative support for this study. They do not represent a random sample of the Pacific region. Schools within each entity were also recruited and selected purposively; they therefore do not represent random samples for each entity. Given the intentional selection process, the pooled data analysis was not designed to produce findings generalizable to the region or to specific entities. Instead, it was designed to examine the effects of the intervention within the entities and schools in the selected sample.

For the primary student outcome analysis, impact was computed as a weighted mean of the effects that were estimated separately by entity. The weighted-average approach was adopted as a pooling method for the student data after initial analyses found that student impact estimates varied across entities. Weights were constructed to take into account the variation in precision of estimates by entity.[48] For the teacher outcome analysis, impacts were estimated based on the

[48] Weights were defined as the inverse of the variance of an effect estimate.

sample combining data from all three entities.[49] Potential differences in outcomes across entities were addressed by including fixed entity effects in statistical models (described below). (For more information on data pooling methods, see appendix B.)

Impact estimation method

In theory, under a random assignment design, a simple comparison of the average outcomes for the two research groups could be used to determine whether the program had any statistically significant impacts. In this study, however, a statistical model was used to account for the effects of clustering teachers and students within each school explicitly, enabling the correct estimation of the standard error of the impact estimates. A statistical model was used also to control for baseline characteristics of schools, teachers, and students, as well as blocking: The inclusion of these covariates increased the study's statistical power by improving the precision of the impact estimates and helped remove any chance differences in the baseline characteristics of treatment and control group schools.

The study used a hierarchical linear model as the primary statistical model for estimating the impacts of Pacific CHILD. Two-level models were used in which teachers and students were nested within schools.[50] (For additional information on how the estimation model was specified, see chapter 5; for technical details, see appendix B.)

Several robustness tests were conducted to document the extent to which the impact estimates were sensitive to the chosen analytic approach. Sensitivity analyses included the application of alternative methods for specifying and estimating the statistical models, handling missing data, and scaling outcome data:

- Alternative estimation methods: The restricted maximum likelihood method was used in the estimation of the benchmark hierarchical linear modeling model. As a sensitivity analysis, the model was also estimated using alternative estimation methods, including the maximum likelihood method, feasible generalized least squares based on the Swamy and Arora method, generalized estimating equations, and ordinary least squares with robust standard errors.
- Alternative methods for treating missing covariates: Listwise deletion was used to address missing covariates for the benchmark model of teacher outcomes. As a sensitivity analysis, the benchmark model was refit using the dummy variable adjustment method.
- Alternative specifications of outcome measures: For the benchmark student outcome measure, the equipercentile linking method was used to pool SAT 10 and TerraNova scale scores. As a sensitivity analysis, z-scores were constructed as an alternatively scaled measure and used to estimate the model. For the benchmark teacher knowledge measure, the alternative measure was the total score from the teacher knowledge assessment. As a sensitivity analysis, impact was estimated using total score excluding low-reliability items and a total score measure based on a two-parameter item response theory model. For the benchmark teacher practice, the alternative measure was the average rating. As a sensitivity analysis, impact was estimated using a measure based on a Rasch partial credit rating model.

[49] The weighted-average approach was not used in the teacher outcome analysis because the small size of the sample did not allow reliable estimation of impacts separately by entity.

[50] Data were not available with which to reliably link teachers and students. Therefore, it was not possible to estimate a three-level model in which students are nested within classrooms.

In addition to these robustness tests, the study also tested random coefficient specifications in which the assumption of fixed coefficients was relaxed. Supplemental analyses of additional teacher outcome measures were also conducted in order to provide additional information to investigate the validity of the original measures. Appendix B provides details on and the results of the sensitivity tests.

Handling of data issues

A common problem in any empirical study is missing data. In addressing the missing data problem, the study adopted different approaches for outcome data and covariates.[51] Listwise deletion—in which individuals with missing values on outcome measures are excluded from the analysis—was used to handle missing observations in the outcome data. In applying listwise deletion, the study assumed that data were missing at random and that individuals excluded from the analytic sample because of missing outcome data were not systematically different from individuals with complete observations.

For missing covariate observations for the student impact analyses, the dummy variable adjustment method was applied.[52, 53] For the teacher outcome analyses, missing covariates were addressed by applying listwise deletion. This approach was used with the teacher data because relatively few observations (less than 5 percent) were missing for each covariate and because the study had a relatively small sample size of teachers, which raised concerns about the loss in degrees of freedom associated with using an alternative approach, such as the dummy variable adjustment method. As a robustness check, the study estimated the models for teacher outcomes using the dummy variable adjustment method for all missing observations for covariates.

Another data issue is the risk of crossover between the treatment and control groups and the resulting contamination of the assigned condition. To minimize crossover, the study team asked school and district officials to prevent teachers from moving from treatment to control group schools and vice versa. Pacific CHILD teachers were also asked not to share Pacific CHILD program materials with teachers from control group schools. The research team, however, did not have control over decisions made by teachers and schools. To avoid arbitrary decisions about classifications of teachers who crossed over, the study design planned to make no adjustments in the impact analysis to account for crossovers. During the intervention, no teachers in the impact sample transferred across conditions.

---

[51] The strategies for dealing with missing observations are consistent with the recommendations of Puma et al. (2009). Chapter 3 discusses missing data (nonresponses) among study-eligible teachers and students. This section summarizes the strategies for dealing with the missing data problem.

[52] Previous studies have discouraged the use of dummy variable adjustment, because it can lead to biased estimates and standard errors for the coefficients of the affected variables (see, for example, Allison 2001). However, a recent simulation study of missing data correction methods in randomized control trials showed that the dummy variable adjustment method is one of the most effective and least technically taxing methods for addressing the problem of missing data among explanatory variables with respect to impact estimates (Puma et al. 2009).

[53] For the dummy variable adjustment method, missing covariates were coded as zero (or the sample mean), and a binary indicator for missing observations was included in the estimation.

Exploratory analyses

Exploratory analyses were designed to enhance the usefulness of the study by providing supplemental information on the findings and pointing to possible directions for future research. The following topics were explored:

- Program impacts on student and teacher outcomes in Hawai‘i
- Program impacts on student outcomes in non-Hawai‘i entities
- Program impacts on subscales of the teacher instructional practice measures
- Moderating effects on program impacts on teacher outcomes by level of experience and education

Chapter 6 discusses the exploratory research questions, analyses, and findings. Appendix C provides technical information on the exploratory analysis models and presents the estimation results.

Chapter 3: Data

This chapter describes the choice of instruments and measures, data collection procedures and schedules, data collection and response rates, and quality assurance procedures used in this study. To assess the impact of Pacific CHILD, the study collected outcome data to measure student achievement in reading comprehension, teacher knowledge, and teacher practice. Implementation data were also collected to document teacher exposure to Pacific CHILD, examine the extent to which the intervention was delivered according to the original design, and monitor whether the randomized study condition was maintained. In addition to outcome and implementation data, the study collected school, student, and teacher background data to describe the sample and to use as covariates in the impact estimation. The chapter includes a timeline of data collection activities (table 3.1), an overview of data types and purposes (table 3.2), and a summary of data completion and response rates (tables 3.3 and 3.4).

Outcome measures and instruments

This section discusses the outcome measures and instruments used to evaluate the impact of Pacific CHILD on student achievement in reading comprehension, teacher knowledge, and teacher practice. Appendixes D and E provide additional information on the instruments, outcome measures, and limitations. Appendix L provides the instruments and protocols.

Student achievement in reading comprehension

To assess whether Pacific CHILD had an impact on student achievement, the study collected student reading comprehension scores on the standardized tests administered in the participating entities. During the study years, the Stanford 10 Achievement Test (SAT 10) was administered in American Samoa and the CNMI, and Hawai‘i the TerraNova, 2nd Edition, was administered in Hawai’i. Both tests have been normed using nationally representative samples, and student scores are comparable to national benchmarks. For grade 5, the reliability of the reading comprehension subtest is .92 for the SAT 10 and .88 for the TerraNova, as measured by the Kuder-Richardson formula 20 coefficient (CTB/McGraw-Hill 2003; Pearson 2004).[54]

Student reading comprehension data were collected for the baseline year in the spring before the intervention began and in the spring of the second year of the intervention (table 3.1).[55] Students completed the tests in March and April at participating schools as part of each school’s annual assessment; reading comprehension scores were collected from participating local educational agencies.[56] Reading comprehension scores of grade 5 students from the second year of the

[54] The Kuder-Richardson formula 20 (KR-20) is a measure of internal consistency reliability for assessments with dichotomous scoring (for example, correct-incorrect). The KR-20 reliability estimate is affected by the range and difficulty of items, the spread in test-taker scores, and the length of assessment.

[55] Local educational agencies provided the student background data discussed in this chapter under the section on school, student, and teacher background data, along with student reading comprehension scores.

[56] In the CNMI and Hawai‘i, standardized tests were administered to grade 5 students each year during the study period as part of the regular assessment schedule for elementary school students. In American Samoa, the SAT 10 was not administered to grade 5 students during the study period. Instead, the American Samoa Department of Education (ASDOE) tested grade 4 students as part of the regular elementary school assessment. The study team requested that

intervention were used in the impact analysis. Of the 3,078 grade 5 students enrolled at the study schools during the spring of the second year of the intervention, 3,052 completed the reading comprehension subtest of either the SAT 10 or TerraNova (99 percent of students enrolled at study schools).[57] Student reading comprehension scores from the baseline year were used to check for baseline equivalence across the conditions at the time of random assignment and used as school-level covariates in the impact estimation.

For the impact analysis, the SAT 10 and TerraNova scores were pooled by converting student reading comprehension scores from the TerraNova to estimated SAT 10 equivalents (see appendix E). Estimated equivalent scores were calculated by using equipercentile methods to link scale scores with the same percentile rank, following the published national norming tables from Pearson and CTB/McGraw-Hill (Kolen and Brennan 2004). For example, a grade 5 student who scored 657 (50th percentile) on the TerraNova would receive a corresponding 50th percentile score of 643 on the SAT 10. By employing a linking approach as the primary method for aggregating the student data, the study retained scale scores as outcome measures. This approach avoided standardizing, recentering scores, and reducing variation across entities. To check the impact analysis results based on the scaled scores, the study created *z*-scores by standardizing student scores to have a mean of 0 and a variance of 1 within each entity-grade subgroup. Appendix E provides details on aggregation using linking and *z*-scores.

ASDOE administer the SAT 10 to grade 5 students during the second year of the study period. ASDOE met the request, administering the SAT 10 to grade 5 students at the same time the grade 4 students were tested, following the same protocols used for grade 4 students. For the confirmatory impact analyses, the school average for grade 5 test results was used as the baseline outcome measure in the CNMI and Hawai‘i, and the school average for grade 4 test results was used as the baseline outcome measure in American Samoa.

[57] The size of the student impact sample was estimated based on school-level data provided by study schools and local educational agencies. Because the student outcome and background data were available only for students who completed the tests, it was not possible to calculate actual response rates for the student assessment. Instead, data completion rates were estimated by dividing the number of student reading comprehension scores by the estimated student impact sample size.

Table 3.1 Data collection schedule for outcome, background, and implementation data

Data collection during first year	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
Student assessment and background data							—	—	—			
Teacher knowledge assessment							—	—	—	—		
Classroom observations						—	—	—	—			
Teacher background survey						—	—	—	—	—		
Principal survey							—	—	—	—		
Teacher focus groups					—	—		—	—			
Program staff interviews										—	—	—
Professional development observations	—	—	—	—	—	—	—	—	—	—	—	—
Participation data and attendance records	—	—	—	—	—	—	—	—	—			—
Data collection during second year	**Sep**	**Oct**	**Nov**	**Dec**	**Jan**	**Feb**	**Mar**	**Apr**	**May**	**Jun**	**Jul**	**Aug**
Student assessment and background data							—	—	—			
Teacher knowledge assessment						—	—	—	—			
Classroom observations					—	—	—	—	—			
Teacher background survey					—	—	—	—	—			
Principal survey							—	—	—	—		
Teacher focus groups						—	—	—	—			
Program staff interviews	—	—							—	—	—	—
Professional development observations	—	—	—		—	—	—	—	—	—	—	
Participation data and attendance records	—	—	—	—	—	—	—	—	—			—

Note: Baseline refers to the 2006/07 school year in American Samoa and the CNMI and the 2007/08 school year in Hawai‘i . First year and second year refer to the 2007/08 and 2008/09 school years in American Samoa and the CNMI and to the 2008/09 and 2009/10 school years in Hawai‘i . The data collection schedule was based on the school calendars in the first and second year and reflects changes in the school calendars from year to year. In the first year, the spring term in Hawai‘i began in late January and ended in early June; in the second year, the spring term began in early January and ended in late May, which affected the data collection schedule. Field staff conducted classroom observations and administered the teacher knowledge assessment in both intervention and control group schools in all entities throughout the data collection period to ensure balanced data collection across conditions and entities. Student data collection schedules show the months in which entities administered the standardized assessments. The research team obtained these test scores from entities after they became available the following school year.

Source: Authors’ summary of data collection activities.

Teacher knowledge

To evaluate the impact of Pacific CHILD on teacher knowledge, the study developed a teacher knowledge assessment consisting of 40 multiple-choice items.[58] Items on the teacher knowledge assessment were developed by the study team and not drawn from existing instruments. The teacher knowledge assessment covers general pedagogical concepts, including both concepts

[58] During data collection in the field, the teacher knowledge assessment was referred to as the teacher impact survey.

emphasized in Pacific CHILD and concepts not emphasized in Pacific CHILD. Items reflecting concepts from Pacific CHILD were developed from the Pacific CHILD manuals and the general pedagogical literature on which Pacific CHILD is based, including research on differentiated instruction, interactive tasks, word parts, and vocabulary.[59] Items that reflect concepts not specifically emphasized in Pacific CHILD were developed from theories and practices covered in undergraduate and graduate coursework, including research on scaffolding, additive bilingualism, and comprehensible input. The study team expected that well-trained teachers in both treatment and control group schools would score well on items that covered concepts emphasized in Pacific CHILD and items that covered concepts not emphasized in Pacific CHILD. Specific content areas covered by the teacher knowledge assessment included English reading instruction, second language acquisition theory, English as a second language instructional methodologies, theories of cognition, principles of lesson planning, assessment and grouping strategies, and scaffolding techniques. Appendix D provides further information on the development and properties of the teacher knowledge survey.

The data used to assess the impact of Pacific CHILD on teacher knowledge were collected between February and May of the second year of the intervention. To ensure balanced data collection across conditions and entities, field staff administered the teacher knowledge assessment to both intervention and control group schools in all entities throughout the data collection period. Teachers were given one hour to complete the assessment in the presence of a field researcher. Of the 236 teachers in the impact sample, 197 (83 percent) completed the teacher knowledge survey.

For the impact analysis, total score (the number of correctly answered questions) was used as the benchmark measure of teacher knowledge. The reliability of the teacher knowledge data collected at the end of the second year was .78, as measured by the KR-20 coefficient. As sensitivity analyses, the study used an item response theory model to create teacher ability measures from teacher responses on the knowledge assessment. An additional alternative total score was created by excluding two items that did not contribute to the overall reliability of the measure. An additional sensitivity analysis was conducted by removing the definitional items on the assessment and examining only items that tested the application of teaching practices. Appendix B provides the results of the sensitivity analyses. Appendix D provides details on the construction of the teacher knowledge measure.

Teacher practice

To assess whether Pacific CHILD had an impact on teacher practice, the study team observed classrooms using a modified version of the Sheltered Instruction Observation Protocol (SIOP®) (Echevarria, Vogt, and Short 2007). Research on the SIOP shows that observers can be trained to use the instrument consistently and that it can be used reliably to evaluate teachers and monitor changes in teacher practice over time (Guarino et al. 2001; Echevarria and Short 2004; Echevarria, Vogt, and Short 2007).

The SIOP consists of 30 items, scored using a five-point Likert scale from zero to four. The 30 items cover 8 areas of instruction across three main dimensions: preparation; instruction

[59] As explained in chapter 1, the content of Pacific CHILD is not unique to Pacific CHILD but is based on accepted research-based reading comprehension and instructional strategies covered in undergraduate and graduate coursework and teacher professional development activities.

(including building background, comprehensible input, strategies, interaction, practice/application, and effectiveness of lesson delivery); and lesson review and evaluation. In addition to the original 30 items, 8 items were added to ensure that the observation protocol captured data for assessing the reading comprehension and instructional strategies emphasized by Pacific CHILD. These eight items were selected from a pool of additional items developed based on a review of the Pacific CHILD training materials and the observation protocol used by program staff during observations of teacher lessons. After elimination of new items that overlapped items in the original SIOP protocol, eight additional items remained. Although designed to measure strategies emphasized in Pacific CHILD, the additional eight items capture general instructional practices that any teacher could use in the classroom.[60] The additional items include focusing on word parts instruction, providing cognitively rich environments, and differentiating instruction. Appendix B provides the analysis excluding the additional eight items. Chapter 6 presents the analyses of subscales for different content areas.

Classroom observations using the modified SIOP were conducted during the second half of the year in both years of the intervention. Trained research staff observed both intervention and teachers at control group schools in all entities throughout the data collection period to ensure balanced data collection across conditions and entities. Before collecting observation data in the field, all observers completed a week-long observation training. They met the minimum interrater reliability thresholds of .80 for at least two of three practice rating sessions that were based on video and in-field observations (interrater reliability was defined as the percentage of the rated items that were in agreement with anchor ratings). During the data collection period, observers completed additional rounds of video checks with the same required threshold of .80, to ensure consistent use of the observation protocol. The average interrater reliability of observers during the second year was .91. Classroom observations were conducted with 198 of the 236 teachers in the impact sample (84 percent).

Classroom observers were not blind to the assignment condition of schools. Their knowledge that a teacher was from a treatment or control group school could have systematically influenced how they rated the teacher, potentially leading to outcome ascertainment bias. To minimize such bias, the observation protocol emphasized the application of specific criteria for rating and required observers to record an observation (or the lack of an observation) of specific activities to justify any high or low ratings. Trainers of observers also warned them of potential ascertainment bias and did not provide details of the intervention or disclose any particular pedagogical approaches or techniques associated with the intervention.

The average SIOP score, adjusted for items that were rated not applicable, was used as the outcome measure of teacher practice in the impact analysis. Of the 30 items on the original SIOP, four items have a not applicable category. After review, the study excluded one of these items in constructing the teacher practice measure, because it was not interpreted and scored consistently across raters and entities. The benchmark outcome measure for teacher practice was thus based on 37 items (29 from the original SIOP and 8 developed by the study team). The

[60] The Pacific CHILD program is based on findings from research on the effectiveness of different reading and instructional strategies. Strategies such as differentiated instruction and focusing on word parts are not specific to Pacific CHILD. These strategies reflect existing and accepted research-based reading comprehension and instructional strategies from the field.

internal consistency reliability of the observation data used to construct the teacher practice measure for the impact analysis was .92, as measured using Cronbach's alpha.[61]

A Rasch model was used to construct an alternative teacher practice outcome measure based on the modified SIOP scores. This measure was used to check the sensitivity of the impact analysis results based on the average score. An additional sensitivity analysis that removed the observational items developed for the study was also conducted. Appendix B provides information on the results of the sensitivity analysis. Appendix D provides additional information on the construction of the outcome measure for teacher practice.

School, student, and teacher background data

In addition to outcome measures, the study collected school- and individual-level background data, in order to check equivalence across conditions at random assignment, describe the samples, and construct covariates for the impact analysis estimation.

School background data

School data from the Common Core of Data (CCD) were used to establish baseline equivalence and increase the precision of estimations in the impact analysis (U.S. Department of Education 2007, 2008). Key variables in the CCD data files included student-teacher ratios, the proportion of racial/ethnic minority students, and the number of students eligible for free or reduced-price lunch. In American Samoa and the CNMI, where implementation began at the beginning of the 2007/08 school year, CCD data from the 2006/07 school year were used as baseline data. In Hawai'i, where implementation began at the beginning of the 2008/09 school year, CCD data from the 2007/08 school year were used as baseline data.

Principal survey

The principal survey was designed to gather descriptive information about the study schools and participants and to serve as covariates in the impact analysis. The principal survey consisted of questions about the schools' grade 4 and grade 5 student and teacher populations, state- and district-level professional development policies, and district- and school-level challenges teachers face. Data from the survey were used to estimate the number of students and teachers in grades 4 and 5 each year of the study. Study principals completed the survey between March and June of the first and second years of the intervention. Principals from all 45 schools completed the principal survey in the first year, the second year, or both years of the intervention[62]

Student background data

After random assignment, data on individual students were collected to use as covariates in the impact estimation. These data included demographic information, including gender, race/ethnicity, and special education needs status for all three entities. For Hawai'i, data were also

[61] Cronbach's alpha is a measure of internal consistency reliability for scales with polytomous or continuous data (for example, Likert scale data).

[62] Exact response rates for the principal survey cannot be reported due to disclosure risk.

collected on eligibility for free or reduced-price lunch and English language learner status. Local educational agencies provided background student data along with reading comprehension scores.

Teacher background data

After random assignment, individual-level teacher background data were collected, including information on gender, race/ethnicity, years of experience, and education level, using the teacher background survey administered to teachers in study schools. Data from the survey were used to assess differences in teacher characteristics across treatment conditions and as covariates in the impact analysis. Teachers at study schools completed the background survey between February and May of the first and second years of the intervention. Data collected toward the end of the second year of the intervention were used to describe the teacher impact sample and as covariates in the teacher-level impact analysis.

The teacher background survey included questions about years of experience, education, credential and certification status, language proficiency, primary language, gender, and race/ethnicity. Of the 236 teachers in the impact sample, 196 (84 percent) completed the background survey toward the end of the second year of the intervention.

Implementation data

The following sections discuss the data collection instruments used to document the study conditions and study sample (table 3.2), examine the extent to which schools and teachers in the study sample received the intervention, and assess the extent to which the intervention was implemented according to the original design.

Participation data and attendance records

Participation data, including attendance logs, were collected from program staff to examine the extent to which teachers in the study sample received the intervention as intended. Pacific CHILD trainers completed monthly participation logs to document the content and duration of each interaction they had with each teacher outside of the annual institutes and mini-institutes. Sign-in sheets from each annual institute and mini-institute were also collected. Participation data were used to estimate the amount of intervention each teacher received (dosage) and the proportion of time they devoted to specific activities and topics during the implementation period. Program staff submitted Pacific CHILD participation logs each month throughout the first and second years of the intervention, as well as attendance records from each annual institute and mini-institute.

Observations of Pacific CHILD professional development activities

Pacific CHILD activity observation protocols and observation data collection forms were designed to record the extent to which the intervention was implemented according to the original design. Trained research staff completed the professional development activity observations. The protocols required observers to compare the designed activities as outlined in the Pacific CHILD professional development manual with the activities they observed. The

protocols also required observers to record the implementation of each professional development activity observed with transcriptive notes and structured summaries.

The professional development observation protocol included prespecified criteria for assessing the fidelity of implementation to the original design. These criteria were developed from the Pacific CHILD professional development manual and interviews with program developers. Observations of Pacific CHILD activities were conducted throughout the first and second years of the intervention. The field research team observed the annual institute and mini-institutes at each site and a sample of the structured learning teams at each treatment group school.[63]

Interviews with program staff

The study team conducted telephone interviews with program developers, project advisors, and trainers to collect information on the study conditions and to determine the extent to which the intervention was implemented according to the original design. Interviews were conducted using a semistructured interview protocol to gather information on the background and training of the trainers, the delivery of the intervention, teacher- and school-level implementation objectives that were met and not met, and external unexpected events and circumstances encountered during implementation. These interviews were conducted May--October during the first and second years of the intervention.

Teacher focus groups and group interviews

The study team conducted focus groups with teachers in treatment group schools to examine the extent to which teachers received the intervention as intended. The semistructured focus group protocol contained open-ended questions designed to elicit information about teachers' experiences of the program and the training they received, satisfaction with and commitment to the intervention, and issues they may have encountered applying what they learned from Pacific CHILD in the classroom. Focus groups with treatment group teachers were conducted January–May during the first year of the intervention and February–May during the second year of the intervention. All teachers who participated in the intervention were invited to participate in the focus groups. Each year, 64 teachers participated in focus groups or group interviews.[64]

---

[63] Observations of mini-institutes included a sample of school-based classroom observations and lesson demonstrations. The second mini-institute in American Samoa was not observed because of a schedule change. Institutes were conducted at the entity level in American Samoa and the CNMI and at two to five locations in Hawai‘i.

[64] Seventy-six teachers were invited to participate in the focus groups the first year. Sixty-nine teachers were invited to participate in the focus groups the second year.

Table 3.2 Summary of data types and purposes

Type of data/instrument	Primary purpose
Outcome measures and instruments	
Stanford 10 and TerraNova	• Determine impact of Pacific CHILD on student achievement in reading
Teacher knowledge assessment	• Determine impact of Pacific CHILD on teacher knowledge
Modified Sheltered Instruction Observation Protocol	• Determine impact of Pacific CHILD on teacher practice
School, student, and teacher background data	
Common Core of Data	• Establish test equivalence across conditions at baseline • Serve as school-level baseline covariates in impact analysis
Student records	• Serve as student-level covariates in impact analysis
Teacher background survey	• Serve as teacher-level covariates in impact analysis
Principal survey	• Serve as school-level covariates in impact analysis • Provide descriptive information on study schools and teachers
Implementation measures and instruments	
Teacher focus groups and interviews (treatment group schools only)	• Determine extent to which teachers received intervention as intended • Provide descriptive information about teachers' experiences with intervention
Professional development observations (treatment group schools only)	• Determine extent to which intervention was implemented with fidelity
Program staff interviews	• Provide information on study conditions • Determine extent to which intervention was implemented with fidelity
Participation data and attendance records (treatment group schools only)	• Determine amount of training each teacher received and proportion of time devoted to specific activities and topics during implementation period

Source: Authors.

Data collection rates, response rates, and quality assurance

Quality assurance and data monitoring procedures were implemented throughout the study. This section discusses data collection rates, response rates, and data quality assurance for the outcome and background data. Appendix K presents additional information on data collection procedures and data quality assurance for the implementation data.

Data completion and response rates

Data completion rates are defined as the proportion of all study-eligible individuals who completed a data collection activity. Data response rates are defined as the proportion of all

consented individuals who completed a data collection activity.[65] Data completion rates were used to check the extent of overall sample attrition; response rates were used to check the results of the fielding of data collection instruments. These rates are summarized in tables 3.3 (for the overall sample) and 3.4 (by the assignment condition)

For students, the data collection process did not require obtaining consent from individuals or contacting them about research activities. Therefore, response rates are not applicable, and only completion rates are reported. For teachers, the completion rate is the number of teachers who completed data collection activities as a percentage of all teachers who could have participated in the study, including teachers who declined to participate in the study and were not contacted about data collection activities. The response rate is the number of teachers who completed data collection activities as a percentage of all teachers who consented to participate in the study and were contacted about research activities.

Table 3.3 Completion and response rates for students and teachers for data collected to measure impact of Pacific CHILD

Data type	Completion rate (percent of all eligible)	Response rate (percent of consented)
Student data		
Student reading comprehension	99.2	Na
Teacher data[a]		
Teacher knowledge assessment	83.5	98.5
Classroom observation	83.9	99.0
Teacher background survey	83.1	98.0

na is not applicable. For students, the data collection process did not require obtaining consent from individuals and contacting individuals about research activities. Therefore, response rates are not applicable, and only completion rates are reported.

a. For teachers, the completion rate is the number of teachers who completed data collection activities as a percentage of all teachers who could have participated in the study, including teachers who declined to participate in the study and were not contacted about data collection activities. The response rate is the number of teachers who completed data collection activities as a percentage of all teachers who consented to participate in the study and were contacted about research activities.

Source: Authors' analysis of student, teacher, and principal participation in data collection activities.

For the student sample, missing outcome data were minimal and balanced across conditions. Student achievement data were available for 3,052 fifth grade students who were tested in the spring of the second year of the intervention. The student sample size was estimated to be 3,078 based on the enrollment information separately collected for the same period. Based on these available data, the overall data completion rate for the student sample was 99.2 percent, with a one percentage point difference between the treatment and control groups (98.7 percent for the treatment group versus 99.7 percent for the control group, $p = .003$).

[65] To collect data from teachers, researchers contacted all study eligible teachers to obtain consent to participate in the study.. Only teachers who consented to be part of the study were contacted for data collection activities.

For the teacher sample, which included 236 grade 4 and grade 5 teachers, 197 (83.5 percent) completed the teacher knowledge assessment and 198 (83.9 percent) responded to the teacher practice observations. For the knowledge assessment, the overall completion rate was 83.5 percent (80.5 percent for the teachers at treatment group schools and 86.4 percent for the control group, $p = .220$). For the teacher practice observations, the overall completion rate was 83.9 percent (81.4 percent for teachers at treatment group schools and 86.4 percent for control group teachers, $p = .290$).

Differences in data completion rates for teacher data were driven predominantly by differences in consent rates across teachers at treatment and control group schools, with teachers at treatment group schools consenting to participate in the study at a lower rate than teachers at control group schools (81.4 percent for teachers at treatment group schools and 88.1 percent of teachers at control group schools, $p = .148$; see chapter 2 for more information). Among teachers who consented to participate, 99.0–100 percent of teachers in treatment group schools and 97.1–98.1 percent of teachers in control group schools participated in the data collection activities.

Table 3.4 Comparison of data completion and response rates between the treatment and control group schools

	Data completion rates			Response rates		
Data source and type	**Treatment schools**	**Control schools**	**Difference**	**Treatment schools**	**Control schools**	**Difference**
Student data						
Student reading comprehension	98.7	99.7	-1.0*	na	na	na
Teacher data[a]						
Teacher knowledge assessment	80.5	86.4	-5.9	99.0	98.1	0.9
Classroom observation	81.4	86.4	-5.0	100.0	98.1	1.9
Teacher background survey	80.5	85.6	-5.1	99.0	97.1	1.9

* Significant at the .05 level (two-tailed test)

na is not applicable. For students, the data collection process did not require obtaining consent from individuals and contacting individuals about research activities. Therefore, response rates are not applicable, and only completion rates are reported.

a. For teachers, the completion rate is the number of teachers who completed data collection activities represented as a percentage of all teachers who could have participated in the study, including teachers who declined to participate in the study and were not contacted about data collection activities. The response rate is the number of teachers who completed data collection activities represented as a percentage of the teachers who consented to participate in the study and were contacted about research activities.

Source: Authors' analysis of student, teacher, and principal participation in data collection activities.

Quality assurance

Student reading comprehension data were transferred directly from each entity and checked for inconsistencies in data records, including missing data, outliers, and other unexpected values. The completeness of student reading comprehension records was checked by triangulating different sources of information on student enrollment at the time of testing, including the

principal survey completed each spring, reports from field consultants, and annual and monthly enrollment figures from the participating departments of education, where available.

For the teacher knowledge assessment, field researchers acted as proctors to ensure that teachers completed the survey within the maximum allotted time (one hour) and under comparable conditions. Classroom observers completed a five-day training and met minimum interrater reliability thresholds at multiple time-points during data collection. To ensure consistency, they used formal protocols to guide their observations and subsequent report writing. All survey and observation data were double entered and discrepancies reconciled, to minimize data entry and processing errors.

To check the quality of the teacher knowledge data collected from the annual teacher impact survey, the study conducted item- and test-level analyses, including item facility, item discrimination, and KR-20 reliability estimates. The quality of the classroom observation data was monitored using Cronbach's alpha to check the internal consistency of the overall instrument and any subscales used in the analysis (see appendix D).

Chapter 4: Implementation of Pacific CHILD

This chapter describes the implementation of Pacific CHILD to provide a context from which to interpret the impact findings presented in chapter 5. The first section describes program activities as designed and reported in the professional development manual developed by REL Pacific. The second and third sections compare the intervention as designed with the actual implementation of Pacific CHILD within the impact sample. Appendix F documents the adaptations made to the original Pacific CHILD design during the implementation period and describes the contextual factors related to program implementation.

Description of Pacific CHILD activities as designed

The Pacific CHILD program was designed as a two-year, school-based intervention that provides sustained, year-round instructional support to participating grade 4 and grade 5 teachers of English language arts. The program includes the following key professional development activities:

- Annual institutes
- Mini institutes
- Year-round instructional support including lesson demonstrations, classroom observations, and peer groups called structured learning teams.

Annual institutes

Pacific CHILD begins each year of the two-year intervention with a 10-day summer annual institute that consists of five days of workshop-style professional development and five days of in-school practice opportunities with students at local elementary schools.[66] The first week of the first annual institute provides an overview of the intervention model and introduces the six Pacific CHILD components. The first week also provides background information on English language learner students and the theory of second language acquisition. The second week of the annual institute is devoted to onsite practice, as teachers apply the concepts learned during the first week in classroom settings with grade 4 and grade 5 students. During the second week, teachers observe lesson demonstrations by program staff, plan and deliver lessons to students, and debrief with program staff, individually and in small groups. Teachers also receive feedback from program staff and other teacher participants in structured learning teams.

The goal of the first annual institute is to provide teachers with a strong foundation in the Pacific CHILD components, which they apply during the school year, as outlined in the activity and component schedule in table 4.1. In the second annual institute, the topics are repeated, in order to strengthen teachers' knowledge by reexamining and practicing each of the six Pacific CHILD components.

[66] Program staff recruited students from nonstudy schools for the onsite practice opportunities during the second week of the summer annual institutes. Annual institutes were held during the summer (June, July, or August) before each study year. The dates were determined based on program staff and teacher availability.

Table 4.1 Sequence of intervention activities and components

Pacific CHILD components	Year 1				Year 2[a]			
	Annual institute	Mini-institutes and structured learning teams			Annual institute	Mini-institutes and structured learning teams		
		Term 1	Term 2	Term 3		Term 4	Term 5	Term 6
Vocabulary								
Word knowledge	✓	✓		✓	✓	✓		✓
Word parts	✓				✓	✓	✓	✓
Text structure								
Sequence[b]	✓	✓			✓		✓	
Compare and contrast	✓		✓		✓	✓		
Cause and effect	✓			✓	✓			✓
Question generation	✓		✓		✓		✓	✓
Differentiated instruction	✓			✓	✓			✓
Cognitively rich environment	✓		✓		✓		✓	
Interactive tasks	✓	✓			✓	✓		

a. During the second year, in addition to the foci presented during the Mini Institutes listed above, teachers are expected to apply differentiated instruction, cognitively rich environments, and interactive tasks during Terms 4, 5, and 6.

b. Text features replaced sequence during the first year of implementation in two entities during the 2007/08 school year. See appendix F for additional information on adaptations to the model during implementation.

Source: Authors' summary of Pacific CHILD teachers' manual (2007).

Mini-institutes

During each school year, the annual institute is supplemented and reinforced by three 3-day mini-institutes (one full day and two half-days), each designed to focus on at least three of the six components introduced in the annual institute (see table 4.1). Mini-institutes follow a structure similar to the annual institutes. Day 1 is a day-long workshop that takes place outside the classroom and highlights the components for the term. Teachers return to their schools and classrooms for two half-days on Days 2 and 3. On Day 2, program staff demonstrate exemplary practices in program school classrooms. On Day 3, program staff visit individual teachers in their classrooms, observe lessons, and provide teachers with feedback. Program staff provide pre- and postconferences for the lesson demonstrations and classroom observations. Both Day 2 and Day 3 involve after-school sessions, to support teacher collaboration, collective reflection, and debriefing in structured learning teams.

Year-round instructional support

Pacific CHILD was designed to provide sustained instructional support to teachers through the following three year-round activities: monthly 45-minute lesson demonstrations, twice-monthly 75-minute classroom observations, and weekly 30-minute structured learning team meetings. These "year-round activities" are designed to increase teachers' understanding of the components covered in the institutes and provide teachers with ongoing support as they implement Pacific CHILD in their classrooms.[67] In addition to the annual institute and mini-institutes, each teacher participating in Pacific CHILD is expected to receive a minimum of four hours of year-round support from program staff every month of the school year (36 hours a year). The following sections describe the year-round instructional support activities as designed.

Lesson demonstrations

Once a month during the school year, program staff model exemplary instructional practices by delivering lessons in treatment group school classrooms that highlight program components. Each lesson demonstration includes a preconference, which allows program staff to discuss the lesson plan and Pacific CHILD components to be targeted during the lesson with teachers. The lesson demonstrations are followed by a postconference, a debriefing session that encourages teachers to question and reflect on their observation of the lesson.

Classroom observations

Twice a month during the school year, program staff observe teachers implementing Pacific CHILD components in their classrooms. Each classroom observation includes a preconference and postconference with teachers to discuss the lesson plans and the implementation of program components and to provide teachers with the opportunity to reflect on their lessons.

School-based structured learning teams

Throughout the two years of the professional development program, structured learning teams meet for 30–45 minutes every week that school is in session. These teams serve as collaborative learning communities; they are designed to facilitate dialogue among teachers and program staff about their experiences with Pacific CHILD. Twice a month, program staff members lead structured learning team meetings, and twice a month teachers meet on their own in structured learning teams, guided by an agenda. Each structured learning team consists of grade 4 and 5 teachers who are participating in the Pacific CHILD program.

Actual exposure to Pacific CHILD

This section compares the intervention design described in the previous section to the actual implementation of Pacific CHILD by examining the extent to which teachers at treatment group schools in the impact sample were exposed to the prescribed intervention. The study is limited in

[67] Year-round instructional support activities occur throughout the school year and do not include the summer months.

its ability to measure exposure to Pacific CHILD by students in the impact sample.[68] Appendix G identifies the challenges associated with measuring student exposure and describes approaches for examining exposure by students. The rest of this section is therefore devoted to teachers' exposure to Pacific CHILD.

The treatment impact sample included grade 4 and grade 5 English language arts teachers who were teaching at the treatment group schools during outcome data collection in the spring of the second year of the intervention.[69] This sample is thus a cross-section of the teacher population at the end of the two-year intervention rather than a sample of teachers who could have attended each Pacific CHILD activity.[70]

The levels of exposure to the intervention were measured by comparing the prescribed intervention activities and participation levels listed in table 4.2 to the actual exposure experienced by the teachers in the impact sample.[71] These participation levels are based on prescriptions for individual teachers rather than for all grade 4 and grade 5 teachers in a school at the end of a two-year program (that is, there is no prescribed standard of participation for teachers at the school level). The total intervention, as designed, includes 42.2 days of professional development (295 hours) over the course of two years.

---

[68] As explained in chapter 2, the student impact sample consisted of grade 5 students enrolled at study schools at the end of the two-year implementation. The discussion here concerns the treatment group in the student impact sample.

[69] The impact sample is the study sample used in the impact analysis presented in chapter 5. As a result, teachers in the impact sample are the primary focus of this chapter. Exposure by participating teachers is presented in the text and in appendix H to provide readers with information on the levels of exposure of teachers who attended intervention activities (that is, excluding teachers with zero participation). Appendix I provides information on exposure by teachers who could have attended all of the annual institutes and mini-institutes (that is, teachers who were present at the treatment group schools during the entire two-year intervention period).

[70] Because the impact sample included only grade 4 and 5 English language arts teachers who were present in the spring of the second year of the intervention, the analysis of exposure to Pacific CHILD presented in this chapter includes some teachers who were not present at the study schools and could not have participated in Pacific CHILD activities during the first year of the intervention. Of the 118 teachers in the impact sample, 98 were present at the study schools during the first annual institute; 100 teachers were present at the study schools for all other Pacific CHILD activities during the first year. Some teachers who participated in the intervention activities in the first year left the study schools after the first year and were not included in the analysis of exposure to Pacific CHILD.

[71] Data on participation in annual institutes and mini-institute were obtained from teacher sign-in sheets and program staff monthly reports. Participation rates in year-round support activities were determined from daily activity logs submitted monthly by program staff during the school year. Attendance at the twice monthly teacher-only structured learning team meetings was not recorded in the participation data and therefore not included in the total number of professional development delivery hours.

Table 4.2 Prescribed annual exposure to Pacific CHILD

Activity	Prescribed participation
Institutes	
Annual institutes (1) ▪ Five days of workshop-style professional development ▪ Five days of in-school practice opportunities with students	10 days a year
Mini-institutes (3)[a] ▪ One day of off-site lecture and small group work ▪ Two half days of in-school demonstration lessons, classroom observations, and program staff–led debriefing sessions	6 days (2 days x 3 mini-institutes) a year
Total institute days	16 days a year
Year-round activities	
Structured learning team meetings ▪ Program staff– led meetings twice a month ▪ Teacher-only meetings twice a month	Weekly (30 minutes)
Lesson demonstrations	Monthly (45 minutes)
Classroom observations	Twice a month (75 minutes)
Total year-round support provided to each teacher	5.1 days (4 hours a month)[b]
Total yearly prescribed intervention	21.1 days (148 hours)[b]
Total prescribed intervention (over two years)	42.2 days (295 hours) [b]

a. Mini-institutes consist of one full day and two half-days. For analysis purposes, the mini-institute hours were combined for a total of two full days of prescribed exposure.

b. Seven-hour days were used to convert activities measured in hours to days.

Source: Authors' summary of Pacific CHILD teachers' manual (Pacific Resources for Education and Learning 2007).

Exposure to annual institutes

Sixty-two of the 118 teachers in the treatment group (52.2 percent) attended the first annual institute (table 4.3).[72] Teachers at treatment group schools, including teachers who did not participate in Pacific CHILD, were exposed to an average of 4.2 days (42.0 percent of the prescribed 10 days) of the first annual institute. Teachers who participated in the first annual institute attended an average of 8.0 days, with 45 of the 62 participating teachers attending eight or more days of the Institute.[73]

Forty-nine of the 118 teachers in the treatment group (41.5 percent) attended the second annual institute. Teachers at treatment group schools, including teachers who did not participate in the program, were exposed to an average of 3.3 days (33 percent of the prescribed 10 days) of the second annual institute. Teachers who participated in the second annual institute attended an average of 8.0 days, with 36 of the 49 participating teachers attending eight or more days.

[73] See appendix H for the average days of attendance among treatment group teachers who attended intervention activities.

Table 4.3 Attendance at Pacific CHILD annual institutes by teachers at treatment group schools in impact sample

Annual institute	Number of attendees	Attendance rate (percent)[a]	Number of attendees by days of attendance				Prescribed days of attendance	Average days attended	Percentage of prescribed intervention received (percent)[b]
			0	1–4	5–7	8–10			
Annual institute 1	62	52.5	56	9	8	45	10	4.2	42.0
Annual institute 2	49	41.5	69	8	5	36	10	3.3	33.3
Average of annual institutes 1 and 2	56	47.0	na	na	na	na	10	3.8	37.7

na is not applicable.

a. Percentage is of all 118 teachers in the treatment group impact sample, 98 of whom were present at the time of the first institute and 117 of whom present at the time of the second institute. Appendix I shows attendance by teachers at treatment group schools in the impact sample who were present in the treatment group schools during the entire two-year intervention.

b. Average days reported are rounded. Percentages reported are based on unrounded average days.

Source: Authors' analysis of annual institute teacher sign-in sheets.

Exposure to mini-institutes

During each school year, three mini-institutes, designed to supplement and reinforce the content presented during the annual institutes, were held. During both years of implementation, the participation rate for each mini-institute was less than 50 percent (table 4.4). Teachers at treatment group schools, including teachers who did not participate in the program, attended an average of 0.8 days per mini-institute, or 39.6 percent of the prescribed two days (one full day and two half days) of the mini-institutes during the first year. Teachers who attended the mini-institutes averaged 1.8 days of attendance per mini-institute the first year, 91.5 percent of the prescribed two days per institute.[74]

During the second year of implementation, teachers at treatment group schools, including teachers who did not participate in the program, attended an average of 0.8 days per mini-institute, or 40.0 percent of the prescribed two days per mini-institute in the second year. Teachers from the impact sample who attended the mini-institutes averaged 1.8 days of attendance per mini-institute the second year, 90.0 percent of the prescribed two days per institute.[75]

[74] The average days of attendance among treatment group teachers who attended the mini-Institutes was 1.9 days for mini-institute 1, 1.8 days for mini-institute 2, and 1.8 days for mini-institute 3.

[75] The average days of attendance among treatment group teachers who attended the Mini Institutes was 1.8 days for mini-institute 4, 1.8 days for mini-institute 5, and 1.7 days for mini-institute 6.

Table 4.4 Attendance at Pacific CHILD mini-institutes by teachers at treatment group schools in impact sample

Mini-institute	Number of attendees	Attendance rate (percent)[a]	Number of attendees by days of attendance: 0 days	1 day or less	More than 1 day	Prescribed days of attendance	Average days attended	Percentage of prescribed intervention received (percent)[b]
Mini-institute 1	55	46.6	63	5	50	2	.9	43.9
Mini-institute 2	50	42.4	68	8	42	2	.8	38.5
Mini-institute 3	48	40.7	70	9	39	2	.7	36.4
Average of mini-institutes 1–3	51	43.2	na	na	na	2	.8	39.6
Mini-institute 4	53	44.9	65	9	44	2	.8	41.1
Mini-institute 5	53	44.9	65	9	44	2	.8	40.7
Mini-institute 6	53	44.9	65	15	38	2	.8	38.3
Average of mini-institutes 4–6	53	44.9	na	na	na	2	.8	40.0
Average of mini-institutes 1–6	52	44.1	na	na	na	2	.8	40.0

na is not applicable.

Note: Mini-institutes consist of one full day and two half days. For analysis purposes, the days were combined for a total of two full days of prescribed exposure.

a. Of the 118 teachers in the impact sample, 100 teachers were present at the treatment group schools at the time of mini-institute 1 (that is, could have attended); 101 teachers could have attended mini-institutes 2 and 3; 117 teachers could have attended mini-institute 4; 116 teachers could have attended mini-institute 5; and 118 teachers could have attended mini-institute 6. Appendix I shows attendance by teachers in the impact sample who were present in the treatment group schools during the entire two-year intervention period.

b. Average days reported are rounded. Percentages reported are based on unrounded average days.

Source: Authors' analysis of mini-institute teacher sign-in sheets.

Exposure to year-round activities

Each month of the school year, each teacher participating in Pacific CHILD was expected to receive a minimum of four hours of year-round support (lesson demonstrations, classroom observations, and structured learning team meetings). During the first year of implementation, 56 of the 118 teachers in the treatment group in the impact sample (47.5 percent) participated in one or more of the year-round activities (table 4.5). On average, these teachers, including teachers who did not participate in the program, were exposed to 1.1 hours of year-round support from program staff each month (27.3 percent of the recommended 4 hours). Teachers from the impact sample who actually participated in year-round activities during the first year received an average of 2.3 hours of support each month (57.5 percent of the recommended 4 hours).

During the second year of implementation, 58 of the 118 teachers in the treatment group in the impact sample (49.2 percent) participated in one or more year-round activities. On average, these teachers, including teachers who did not participate in the intervention, were exposed to an average of 1.0 hours of year-round intervention support (25.2 percent of the recommended 4 hours). Teachers from the impact sample who actually participated in year-round activities during the second year received an average of 1.9 hours of support each month (47.5 percent of the recommended 4 hours).

Table 4.5 Year-round Pacific CHILD support received per month by teachers at treatment group schools in impact sample

Year	Number of teachers who attended at least one year-round activity	Number of attendees, by hours of attendance			Prescribed hours	Average hours[a]	Percentage of prescribed intervention received[b]
		0 hours	Less than 2 hours	2 or more hours			
First year	56	62	23	33	4	1.1	27.3
Second year	58	60	28	30	4	1.0	25.2
First and second years	57	na	na	na	4	1.0	26.0

na is not applicable.

Note: The average hours of year-round support per month was determined by dividing the total number of year-round support activity hours provided per year by the nine months of the school calendar year. Of the 118 treatment groups teachers, 100 were present at the treatment group schools at the time of the year-round activities the first year, and 118 were present at the treatment group schools at the time of the year-round activities the second year. Appendix I provides information on receipt of year-round support activities by teachers at treatment group schools in the impact sample who were present in the treatment group schools during the entire two-year intervention period.

a. During the first year, teachers received an average of 0.35 hours of program staff–led structured learning teams meetings, 0.27 hours of demonstrations, and 0.46 hours for classroom observations. During the second year, they receive 0.28 hours of program staff–led structured learning teams meetings, 0.25 hours of demonstrations, and 0.47 hours of classroom observations.

b. Average hours are rounded. Percentages reported are based on unrounded average hours.

Source: Authors' analysis of year-round support monthly activity logs.

Exposure to intervention: summary

Overall, teachers at treatment group schools from the impact sample were not exposed to Pacific CHILD at the levels prescribed by the intervention (table 4.6). Over the course of the two-year intervention, the average number of days of annual institute and mini-institute attendance by teachers at treatment group schools in the impact sample, including those who did not participate in the program, was less than half (33.3–43.9 percent) the intended number of days per institute (see tables 4.3 and 4.4). Average total exposure to year-round activities was 26.0 percent of the prescribed hours across the two years of the intervention (see table 4.5).

Table 4.6 Summary of exposure to Pacific CHILD by teachers in impact sample

(days, except where otherwise indicated)

Year/activity	Prescribed exposure	Average exposure of teachers in treatment group	Percentage of prescribed intervention received[a]
First year			
Annual institute 1 and mini-institutes 1–3	16	6.6	41.1
Year-round support activities	5.1 (36 hours)	1.4 (9.8 hours)	27.3
Subtotal	21.1	8.0	38.9
Second year			
Annual institute 2 and mini-institutes 4–6	16	5.7	35.6
Year-round support activities	5.1 (36 hours)	1.3 (9.0 hours)	25.2
Subtotal	21.1	7.0	33.1
Total (both years)	42.2 (295 hours)	15.0 (105 hours)	35.6

a. Average days and hours reported are rounded. Percentages reported are based on unrounded average days and hours.

Seven-hour days were used to convert activities measured in hours to days to calculate total year-round support received per year and total hours of prescribed intervention.

Source: Authors' analysis of annual institute and mini-institute teacher sign-in sheets and program staff's monthly activity logs.

In the first year of implementation, teachers at treatment group schools, including those who did not participate in the program, were exposed to an average of 8.0 of the prescribed 21.1 days of intervention activities (38.9 percent of the prescribed intervention). In the second year, exposure decreased to an average of 7.0 days (33.1 percent of the prescribed intervention). Over the course of the two years of the intervention, teachers at treatment group schools in the impact sample, including teachers who did not participate in the program, were exposed to an average of 15.0 days of the prescribed 42.2 total days (35.6 percent of the prescribed intervention). Teachers who participated in the program during both the first and second years of the intervention averaged

31.4 days of exposure (74.3 percent of the prescribed intervention).[76] Table 4.6 provides a summary of exposure to Pacific CHILD by teachers in the impact sample. Appendix H provides the average days of attendance of teachers at treatment group schools who attended intervention activities. Appendix J provides additional information on exposure patterns by entity.

Table 4.7 Exposure to Pacific CHILD by teachers at treatment group schools in impact sample, by level of participation

Level of participation	Number of teachers	Average total exposure per teacher,[a] in days	Percentage of prescribed intervention received
Teachers at treatment group schools in impact sample	118	15.0 (105.0 hours)	35.6
Did not participate in any activity	50	0	0
Participated in at least one activity	68	26.0 (182.2 hours)	61.8
Attended activities in first year only	10	3.5 (24.7 hours)	8.4
Attended activities in second year only	4	10.1 (70.6 hours)	23.9
Attended activities in both first and second years	54	31.4 (219.6 hours)	74.4
Attended all institutes (annual institute and mini-institutes)	36	35.0 (244.7 hours)	83.0
Missed one institute[b]	10	26.6 (186.4 hours)	63.2
Missed two or more institutes[c]	8	21.2 (148.2 hours)	50.2

a. Average days and hours reported are rounded. Percentages reported are based on unrounded average days and hours. Seven-hour days were used to convert activities measured in hours to days to calculate total year-round support received per year and total hours of prescribed intervention.

b Six of the 10 teachers who missed an institute missed an annual institute.

c. Fewer than three of the eight teachers who missed an institute missed one or both of the annual institutes.

Source: Authors' analysis of annual institute and mini-institute teacher sign-in sheets and program staff monthly activity logs.

The difference between the prescribed levels of participation and the actual proportion of intervention received in the treatment group largely reflects the fact that some teachers did not participate in any program activities (table 4.7). Of the 118 teachers at treatment group schools in the impact sample, 50 teachers (42.4 percent) did not complete any Pacific CHILD program activities during the two-year program and were therefore not exposed to any of the

[76] The average exposure among teachers who participated in Pacific CHILD includes the 54 teachers who attended intervention activities during the first and second years of the intervention.

intervention.[77] Sixty-eight teachers (57.6 percent) participated in at least one Pacific CHILD activity during the course of the two-year intervention. Of the 68 teachers who participated in any Pacific CHILD activity, 46 (67.7 percent) received year-round instructional support and missed no more than one institute during the two-year intervention.

Actual delivery of Pacific CHILD activities

This section discusses the extent to which planned intervention activities were delivered by program staff as intended. It examines average activity duration and the frequency of activity delivery.[78] In contrast to the previous section, which focused on teacher participation in each activity, this section focuses on the delivery of each intervention activity regardless of how many teachers participated.

Duration of annual institutes and mini-institutes

Pacific CHILD was designed to provide teachers with two 10-day annual institutes and six 3-day mini-institutes. Program staff delivered a total of 10 annual institutes across the three entities during the two-year intervention. Eight of the 10 annual institutes consisted of the planned duration of 10 days across two weeks.[79] Two of the 10 annual institutes lasted nine days.[80] Mini-institutes were designed to take place over the course of three consecutive days. In practice, mini-institute delivery was spread across an average of 12.7 calendar days, with a range of 3–23 calendar days (see appendix F).

Table 4.8 Prescribed and actual duration of year-round Pacific CHILD support activities delivered to teachers at treatment group schools in impact sample over two-year intervention

Activity	Prescribed length (minutes)	Actual length (minutes)	Percentage of prescribed intervention delivered[a]
Lesson demonstrations	45	63.8	141.7
Classroom observations	75	69.4	92.5
Program staff–led structured learning teams	30	42.2	140.3

a. Average minutes reported are rounded. Percentages reported are based on unrounded numbers.

Source: Authors' analysis of program staff monthly activity logs.

[77] Of the 50 teachers who did not participate in any activities during the two-year intervention, 15 did not teach in the study schools during the first year of the study.

[78] Participation data, including annual and mini-institute teacher sign-in sheets and program staff year-round support, were used in the analysis presented in this section.

[79] Program staff delivered each annual institute at multiple sites in Hawai'i and at a single site in American Samoa and the CNMI. Program staff also delivered two-day annual institute make-up session in some cases for teachers in the impact sample who could not attend any portion of the second 10-day institute.

[80] Institutes were reduced to nine days due to unexpected circumstances..

Duration of year-round activities

Year-round activities were designed to last 75 minutes for classroom observations, 45 minutes for lesson demonstrations, and 30 minutes for structured learning team meetings. In the first and second years, the average duration of the classroom observation, including preconferences and postconferences, was shorter than the prescribed duration of 75 minutes, with an average of 69.4 minutes (92.5 percent of the prescribed duration). In contrast, lesson demonstrations and structured learning team meetings led by program staff exceeded the prescribed length, averaging 63.8 and 42.2 minutes respectively (141.7 percent and 140.3 percent of the prescribed duration) (see table 4.8).

Frequency of Pacific CHILD activities

The Pacific CHILD program was designed to include a prescribed number of activities each year of the two-year intervention period: one Annual Institute, three Mini Institutes, monthly demonstrations by program staff, twice-monthly classroom observations, and weekly structured learning team meetings, with alternate weeks facilitated by program staff. Analysis of program staff logs indicated that program staff delivered the prescribed number of annual institutes and mini-institutes during the two-year intervention; they did not provide the prescribed frequency of year-round activities, however. The highest proportion of prescribed frequency was for demonstration lessons (52.8 percent), followed by structured learning team meetings led by program staff (46.4 percent) and classroom observations (43.6 percent). Although the average duration of the year-round activities exceeded the designed length (see table 4.8), these activities did not occur at the prescribed frequency. For example, treatment teachers received an average of 15.7 classroom observations over the course of the two-year program rather than the 36 classroom observations prescribed by intervention. This contributed to the lower levels of year-round intervention support per teacher each month. The frequency of intervention activities provided by program staff is summarized in table 4.9.

Table 4.9 Prescribed and actual frequency of Pacific CHILD activities delivered to teachers at treatment group schools in impact sample over two-year intervention

Activity	Prescribed frequency	Actual frequency	Percentage of intervention delivered with prescribed frequency[a]
Annual institute (1 per year)	2	2	100.0
Mini-institutes (3 per year)	6	6	100.0
Lesson demonstrations (1 per month)	18	9.3	52.8
Classroom observations (2 per month)	36	15.7	43.6
Structured learning team meetings led by program staff (2 per month)	36	16.7	46.4

a. Average frequencies reported are rounded. Percentages reported are based on unrounded activity frequency.

Source: Authors' analysis of annual institute and mini-institute teacher sign-in sheets and program staff monthly activity logs.

Summary

On average, teachers at treatment group schools in the impact sample were not exposed to the levels of the intervention as designed. Over the course of two years, the 118 teachers in the impact sample, including teachers who did not participate in the program, were exposed to an average of 15.0 of the 42.2 days of prescribed intervention (35.6 percent of the prescribed intervention). The difference between the prescribed levels of exposure and actual exposure in the treatment group largely reflected the effect of the 50 teachers in the impact sample who did not participate in any Pacific CHILD program activities during the two-year program. Appendix J provides additional information on exposure patterns by entity.

Fidelity to the original intervention delivery design varied across program activities. Program staff provided the prescribed frequency of annual institute and mini-institutes, and the average duration of the year-round activities almost met or exceeded the designed duration (92.5 percent for classroom observations, 141.7 percent for lesson demonstrations, and 140.3 percent for program staff-led structured learning team meetings). Program staff did not deliver the prescribed frequency of year-round activities, however, and not all mini-institutes were delivered over the course of three consecutive calendar days, as designed. Appendix F provides additional information on the primary adaptations made to Pacific CHILD during the implementation period and the contextual factors related to program fidelity and teacher participation.

The analysis in this chapter does not reflect the informal support provided to teachers by program staff beyond the formally prescribed Pacific CHILD activities.[81] It also excludes teacher-led structured learning teams, for which verifiable activity logs were not available. As a result, the analysis may not fully capture the total extent of the intervention delivered to participating teachers and schools.

[81] Examples of informal and spontaneous support provided by program staff include after-hours phone or email consultations on lesson plans and instructional delivery, provision of instructional resources, and in-classroom assistance to students.

Chapter 5: Impact analysis findings

As discussed in chapter 2, This chapter presents the impact findings on students and teachers. Appendix B provides additional information on the statistical modeling, covariates, data pooling approaches, estimation results, and sensitivity analyses.

Impact of Pacific CHILD on students

The primary research question for this study examines whether Pacific CHILD improved student outcomes at the end of the two-year intervention. Student outcomes were measured by the reading comprehension subscores of the Stanford 10 Achievement Test (SAT 10) and TerraNova. TerraNova scores were converted to SAT 10-equivalent scale scores (for information on how student achievement data were pooled across tests, see chapter 3 and appendix E). To describe the base performance among students under the business as usual condition, table 5.1 provides basic univariate statistics of the student outcome measure for the control group.

Table 5.1 Scale scores on reading comprehension component of Stanford 10 Achievement Test of students at control group schools

Entity	Mean score	Standard deviation	Minimum score	Maximum score	Number of observations
All entities combined	634.5	36.9	552	734	1,486
American Samoa	600.7	20.6	552	663	64
Commonwealth of the Northern Mariana Islands	631.0	29.4	552	713	320
Hawai‘i	637.5	38.6	556	734	1,102

Note: Scores are based on reading comprehension assessment data from the Stanford 10 Achievement Test (SAT 10) for American Samoa and the Commonwealth of the Northern Mariana Islands and the TerraNova for Hawai‘i. TerraNova scores were converted to SAT 10–equivalent scores using the methodology described in appendix E.

Source: Authors’ calculations based on student test records collected for this study.

For the student outcome, the overall impact across the three study entities was estimated based on a weighted-average approach conducted in two steps. The first step was to estimate three entity-specific impacts independently using the same set of covariates. The estimation model included the following covariates: student characteristics (gender and special education status); school characteristics (baseline school average scale scores in reading comprehension, baseline school size, baseline student-to-teacher ratio, and baseline percentage of students eligible for free or reduced-price lunch); and assignment block indicators. For each entity, impact was estimated based on the restricted maximum likelihood estimation of a two-level hierarchical linear model. Missing outcome data were handled using listwise deletion.[82]

[82] No covariates were missing for student observations with nonmissing outcome data.

The second step of the estimation was to calculate the overall impact across the study entities by averaging across the three entity-specific estimates. The average was computed, weighting the precision of each entity estimate.[83] The estimates are reported in actual measurement units (scale score points) as well as in effect sizes (table 5.2). The effect size shows the impact estimate in the standardized unit, based on the standard deviation of the outcome measure for the control group.[84]

Table 5.2 Intent-to-treat impact estimates of Stanford 10 Achievement Test scale score for reading comprehension

	Regression-adjusted means						
Entity	**Treatment group**	**Control group**	**Difference (impact)**	**Standard error**	***p*-value**	**Effect size**	**Number of observations (students)**
Overall impact, weighted mean of three entity estimates	634.3	629.0	5.3*	2.19	.017	.244	3,052
Entity-specific estimates used to compute the overall impact							
American Samoa	595.6	598.7	–3.0	6.24	.629	–.146	185
Commonwealth of the Northern Mariana Islands	636.1	624.6	11.5*	5.23	.027	.392	692
Hawai‘i	640.6	635.5	5.1	2.62	.050	.133	2,175

*Significant at the 0.05 level (two-tailed test).

Note: For each entity, regression-adjusted means were computed at the means of the covariates; effect sizes were calculated by dividing the impact estimate by the standard deviation of the control group. Each entity estimation included the following covariates: blocking variables, school-level baseline reading comprehension scale score, school size, student-to-teacher ratio, percentage of students eligible for free and reduced-price lunch, student gender, and student special education status. The overall impact and effect size were computed as weighted means of the single-entity estimates, using the inverse of the variance of the entity-specific impact estimate as weights. The results are based on restricted maximum likelihood estimation.

Source: Authors' analysis based on student records collected for this study.

The results show a statistically significant difference in reading comprehension scores between grade 5 students in treatment group schools and their counterparts in control group schools in the spring of the second year of the intervention. The estimated difference in the SAT 10 reading comprehension scores was 5.3 points (effect size = 0.244), statistically significant at the five

[83] See appendix B for the estimation results from the unconditional model and estimates for the unconditional intraclass correlation.

[84] The effect size was computed as the weighted mean of the entity-specific effect sizes. For each entity, the effect size was estimated by dividing the regression-adjusted mean difference in reading comprehension scores between the control and treatment groups by the standard deviation of the control group mean. The overall effect size is then computed as the mean of the entity effect sizes, applying the same weights used to compute the overall impact estimates in the scale score unit.

percent level (p = .017). As a reference point for the size of the impact, the nominal gain on the SAT 10 reading comprehension subscore between grade 4 and grade 5 students is 12 points at the 50th percentile for a nationally representative sample (Pearson 2004).

Several sensitivity analyses were conducted to assess whether the findings remained robust with respect to the estimation method, the linking method used to construct the student outcome measure, and the covariate slope specification.[85] All alternative estimates from different estimation methods and covariate specifications were inside one another's confidence intervals. As with the benchmark model, when the data across entities were pooled using the weighted-average approach, all sensitivity analyses conducted yielded statistically significant results.[86]

The weighted-mean estimate reported in table 5.2 represents the impact for the overall study sample, pooling the data from the three study entities. The results do not imply that the intervention was equally effective across the study entities. Exploratory analyses suggest that the effectiveness of the intervention may have varied by entity in the study sample (see chapter 6).

Impact of Pacific CHILD on teachers

The secondary research question for this study examines whether Pacific CHILD improved teacher knowledge or instructional practice at the end of the two-year intervention

Teacher knowledge was measured by the total score on a multiple-choice assessment. Teacher practice was measured by the average rating from an observation protocol. To describe the base-level performance of teachers in the absence of the intervention, table 5.3 provides the basic statistics of the teacher outcome measures for the control group.

The teacher outcome analyses were conducted by pooling the data from the three entities into a single combined sample. The weighted-average approach used for the student impact estimation was not applied, because the teacher sample size was not large enough to produce reliable estimates by entity. Potential between-entity variation in the program effects was instead addressed by including entity indicators as covariates in the combined sample analysis.

[85] Alternative estimators for the impact coefficient and variance components included those based on maximum likelihood, feasible generalized least squares with an analysis-of-variance estimator of the covariance matrix, generalized estimating equations, and ordinary least squares regression with robust clustered standard errors. As an alternative method for linking the student test scores across entities, z-scores were used. As an alternative covariate specification, the slope of a covariate was treated as random.

[86] The estimation results may be sensitive to the choice of the approaches used to pool the data from the three entities. The initial student impact analysis was conducted based on the combined sample in which data from the three entities were pooled. The combined sample approach, which accounted for variation across entities by fixed entity effects, did not yield consistent inferences across alternative estimation methods. See appendix B for discussions of these results.

Table 5.3 Outcome measures for teachers at control group schools at baseline

Domain	Instrument	Outcome variable	Mean score	Standard deviation	Minimum score	Maximum score	Number of observations
Teacher knowledge	40-item assessment of pedagogical knowledge with focus on instructional needs of English language learner students	Total score (items scored with 1 if correctly answered, 0 otherwise)	25.3	5.61	8	35	102
Teacher practice[a]	38-item classroom observation measure of instructional quality	Average of 5-point (0–4) scale items, adjusted for items that were not rated	1.89	.560	.63	3.22	102

a. Average score computed based on 37 items, excluding one item that was not rated consistently. See chapter 3 for more information on content areas covered by outcome measures.

Source: Authors' analysis of data collected for this study.

The estimates for teacher outcomes were based on the restricted maximum likelihood estimation of a two-level hierarchical linear model. In addition to entity indicators, the model included the following covariates: assignment block indicators; a set of school-level characteristics (baseline school average scale scores in reading comprehension, school size, student-to-teacher ratio, and percentage of students eligible for free or reduced-price lunch); and a set of teacher-level covariates (grade taught, gender, primary language, total years of teaching at baseline, total years at current school at baseline, highest degree completed, teaching certification status, and race/ethnicity). Missing data for the outcome variables and covariates were handled using listwise deletion.

As with the student outcomes, the impact estimates are reported both as actual measurement units and effect sizes (table 5.4).[87] To answer the secondary research question, the study tested the null hypothesis that there was no impact in either teacher outcome. This hypothesis test involved separate significance tests on the two teacher outcomes, adjusting for multiple comparisons based on the Benjamini-Hochberg procedure. After adjustment of *p*-values for multiple comparisons, the study found statistically significant differences between teachers at treatment and control group schools in both of the teacher outcome measures. The null hypothesis of no impact in either teacher outcome was thus rejected.

Having established that Pacific CHILD had impacts across the domains, the results indicate the following for each teacher outcome. For the teacher knowledge measure, the average total score on a 40-point test was 27.0 points for teachers at treatment group schools and 25.0 for teachers at control group schools. This difference of 1.96 points is statistically significant (effect size = 0.35, $p = .023$, *adjusted* $p = .023$). For the teacher practice measure, the average observation score (on

[87] See appendix B for the estimation results from the unconditional models and for estimates of unconditional intraclass correlations.

a 5-point scale ranging from 0 to 4) was 2.20 for teachers at treatment group schools and 1.85 for teachers at control group schools. This difference of 0.36 is statistically significant (effect size = 0.64, $p = .003$, *adjusted* $p = .006$).

Table 5.4 Intent-to-treat impact estimates on teacher outcomes

	Regression-adjusted means							
Impact measure	**Treatment schools**	**Control schools**	**Difference**	**Standard error**	***p*-value**	**Adjusted *p*-value**	**Effect size**	**Number of observations**
Teacher knowledge: Total knowledge assessment score	27.0	25.0	1.96*	.864	.023	.023	.35	190
Teacher practice: Average classroom observation score	2.20	1.85	.36**	.121	.003	.006	.64	189

*Significant at the .05 level (two-tailed test). **Significant at the .01 level (two-tailed test).

Note: Regression-adjusted means were computed at the means of the covariates. Effect sizes were calculated by dividing the impact estimate by the standard deviation of the control group. Adjusted *p*-values were calculated using the Benjamini-Hochberg procedure to control for false discovery rate for the cross-domain comparisons across teacher outcomes.

Source: Authors' analysis based on the primary data collected for this study.

A series of sensitivity analyses were conducted to assess whether the findings remained robust with respect to the estimation method, the method of handling missing data, and the specification of the outcome measures (see appendix B).[88] These analyses were conducted separately for each teacher outcome measure. For both teacher knowledge and practice, alternative methods of analysis yielded results that were consistent with the findings reported in table 5.4 in both the size and statistical significance of the impact estimates.

To provide additional information on the content validity of teacher outcome measures, the study also estimated impact using alternative knowledge measures based on subcomponents of the knowledge assessment as well as an alternative practice measure based only on the original Sheltered Instruction Observation Protocol items.[89] The effects estimated using alternative measures are statistically significant (see appendix B).

[88] As with the student outcome estimation, alternative estimators were based on maximum likelihood regressions, feasible generalized least squares regressions with an analysis-of-variance estimator of the covariance matrix, generalized estimating equations, and ordinary least squares regressions with robust clustered standard errors. As an alternative method for handling missing data, the dummy variable adjustment was applied. As alternative outcome measures, scores that were scaled differently and specified without underperforming items were tested.

[89] To address concerns that the instrument may have been testing vocabulary rather than pedagogical knowledge, the study estimated the impact of Pacific CHILD on teacher knowledge based on the total score for definitional questions and the total score for nondefinitional questions. The impact remained statistically significant.

Chapter 6: Exploratory analyses

This chapter investigates a set of research questions that explore the impact patterns beyond the confirmatory analyses. It explores impacts by entity, impacts on different dimensions of teacher practice, and impacts on teacher outcomes by experience and education levels. The purpose of these exploratory analyses is to examine empirical patterns of program impacts and generate questions for future investigation.

In contrast to the confirmatory analyses presented in chapter 5, these exploratory analyses do not test specific hypotheses regarding causal relationships between the intervention and outcomes derived from the underlying theoretical model. In addition, in contrast to the models used to estimate the overall impacts reported in chapter 5, the models for the exploratory analysis of impacts by entity incorporate additional control covariates in order to obtain more precise estimates for each entity. This chapter discusses differences in the estimation approach for the confirmatory and exploratory analyses of entity-specific impacts. (Appendix C provides technical details on the analyses.)

Impact of Pacific CHILD on students and teachers in Hawai‘i

Hawai‘i is the largest and most populous entity in the region served by REL Pacific. Its statehood status is also unique in the region. For these reasons, this study was designed with a sufficiently large sample of schools in Hawai‘i to support subgroup analyses in Hawai‘i. Impact analyses specific to Hawai‘i provide additional information for stakeholders and policy audiences in Hawai‘i. They were not intended to produce estimates that are directly comparable to impacts in other entities or the overall pooled impact. This subgroup analysis should therefore be used only to understand the impact patterns within the Hawai‘i subsample.

Impact on students in Hawai‘i

The subgroup analyses for the Hawai‘i subsample explored the following research question, which is parallel to the primary question examined in chapter 5:

- Exploratory question 1: Did grade 5 students at schools in Hawai‘i that were offered Pacific CHILD for two years perform differently on assessments of reading comprehension from grade 5 students at schools in Hawai‘i that were not offered Pacific CHILD?

The study also explored the impacts of the intervention based on an alternatively defined subgroup of students in Hawai‘i that included students who were present in the study schools from the time of random assignment until the end of the intervention. This subsample consisted of students who were in grade 3 at the time of random assignment and grade 5 students near the end of the study. In contrast to the confirmatory impact analysis sample, which focused on all grade 5 students present in study schools toward the end of the intervention, this subgroup analysis examined impacts on a cohort of students tracked over time. The question explored for this subgroup analysis was as follows:

- Exploratory question 2: Did students enrolled for the entire study period at schools in Hawai‘i that were offered Pacific CHILD perform differently on assessments of reading

comprehension conducted after two years of implementation of the intervention from their counterparts at schools in Hawai‘i that were not offered Pacific CHILD and who were enrolled at those schools for the entire study period?

A key difference between the analyses of exploratory questions 1 and 2 is the student samples. Exploratory question 1 examines students who happened to be in the study schools at the end of the intervention period. Exploratory question 2 examines students who were in the study schools from the start through the end of the intervention period. Exploratory question 1 focuses on the impacts on students at schools that were potentially exposed to the intervention. Exploratory question 2 focuses on impacts on students who were potentially exposed to the intervention.

For the confirmatory research questions, the study estimated entity-specific impacts on students as the first step of the weighted-average analysis of the overall impact. This step emphasized that the equivalent model was applied in all entities. In contrast, for the exploratory subgroup analyses in this chapter, the study applied the most appropriate approach for each setting.

Like the confirmatory impact analyses, the Hawai‘i -specific impacts on students were estimated based on a hierarchical linear model, using the same student outcome measure (the Stanford 10 Achievement Test [SAT 10]–equivalent reading comprehension scale scores). The inclusion of additional student-level covariates distinguishes the exploratory analyses from the first step of the confirmatory analysis, in which impacts on students were estimated using only the covariates commonly available across all three entities. In the entity-specific estimation used for the confirmatory analysis, only gender and special education status were included as individual-level demographic covariates. For the exploratory analysis of impacts on students in Hawai‘i, three individual-level covariates were added to the estimation model: English language learner status, race/ethnicity, and eligibility for free or reduced-price lunch. The model also included student-level baseline reading comprehension scores, instead of school-level scores used for the confirmatory analysis. For the school-level baseline covariates, the percentage of English language learner students was included in addition to those included in the confirmatory analysis (school size, student-to-teacher ratio, percentage of students eligible for free or reduced-price lunch, and assignment block indicators). Inclusion of the additional student- and school-level covariates was expected to help explain the variation in the outcome reading comprehension scores across individuals and schools, thereby increasing the precision of estimation of the impacts. (See table C.3 in appendix C for a summary of the covariates used in the subgroup analyses.)

The regression-adjusted estimates of the impact of Pacific CHILD on the reading comprehension scores of students in Hawai‘i were statistically significant for both the sample of all grade 5 students at study schools at the end of the two-year intervention (exploratory question 1) and the sample of grade 5 students who had been at study schools for two years (exploratory question 2) (table 6.1).[90] The model specified to answer exploratory question 1 resulted in an estimated difference in reading comprehension scores of 4.04 points (effect size = 0.10, p = .037). The model specified to answer exploratory question 2 resulted in an estimated impact of 4.03 points (effect size = 0.11, p = .008). These results indicate that after nearly two years of implementation in Hawai‘i, Pacific CHILD had statistically significant impacts on student outcomes.

[90] The model used to analyze exploratory question 1 was fit using the restricted maximum likelihood method. The model used to analyze exploratory question 2 was fit using feasible generalized least squares, because restricted maximum likelihood routines failed to converge.

Table 6.1 Estimated impact of Pacific CHILD on Stanford 10 Achievement Test–equivalent reading comprehension scale scores of grade 5 students in Hawai‘i

Sample (corresponding exploratory research question number)	Regression-adjusted means					Effect size	Number of observations
	Treatment schools	Control schools	Difference	Standard error	*p*-value		
Grade 5 students enrolled at study schools toward end of year 2 (exploratory question 1)	640.0	636.0	4.04*	1.94	.037	.10	2,175
Grade 5 students enrolled at study schools at baseline and toward end of year 2 (exploratory question 2)	640.4	636.4	4.03**	1.51	.008	.11	1,741

SAT 10 is the Stanford 10 Achievement Test.

*Significant at the .05 level (two-tailed test), **significant at the .01 level (two-tailed test).

Note: Regression-adjusted means were computed at the means of the covariates. Effect sizes were calculated by dividing the impact estimate by the standard deviation of the control group. The model for exploratory question 1 was estimated using the restricted maximum likelihood method. The model for exploratory question 2 was estimated using feasible generalized least squares. TerraNova scores were converted to SAT 10 scores using the methodology described in appendix E.

Source: Authors' calculations based on data collected for this study.

Impact on teachers in Hawai‘i

Subgroup analysis of teachers in Hawai‘i were based on the following research questions, which are parallel to the secondary question examined in chapter 5:

- Exploratory question 3: Did grade 4 and grade 5 teachers at schools in Hawai‘i that were offered Pacific CHILD for two years perform differently on assessments of their knowledge of theories and strategies related to effective reading instruction, including English language learner-focused theories and strategies, from teachers at schools in Hawai‘i that were not offered Pacific CHILD?
- Exploratory question 4: Did grade 4 and grade 5 teachers at schools in Hawai‘i that were offered Pacific CHILD for two years perform differently on an assessments of their instructional practices, including English language learner-focused practices, for enhancing student reading comprehension from teachers at schools in Hawai‘i that were not offered Pacific CHILD?

The impacts on teacher outcomes in Hawai‘i were estimated using a hierarchical linear modeling approach. In contrast to the exploratory analyses, the two teacher outcomes were analyzed independently without adjustment for multiple comparisons. The same outcome measures were used in the exploratory subgroup analyses as in the confirmatory analysis (the total score from the teacher knowledge assessment and the average score from teacher practice observations). The exploratory subgroup analyses included the following teacher-level covariates: gender, grade taught, primary language, total years of teaching, years at current school, education, and race/ethnicity (the same set of teacher-level covariates used in the confirmatory pooled estimation, excluding teaching certification status [all teachers in Hawai‘i are certified]). Given the small sample size, a subset of the school-level covariates used in the confirmatory pooled

estimation was included in the estimation to reduce the loss of degrees of freedom. Specifically, the estimation model for the exploratory subgroup analyses for teachers in Hawai‘i included school size at baseline and assignment block indicators.[91] For the teacher practice outcome estimation, the baseline school-level average for student reading comprehension was also included in the model.[92]

Like the exploratory student analysis, the exploratory teacher analysis was based on an estimation model that was different from the model used in the confirmatory analysis. The results presented in table 6.2 are therefore not intended to be compared with the overall pooled impact estimate or with impacts for other entities.

No statistically significant difference was detected between treatment and control groups in the teacher knowledge measure. In contrast, a statistically significant difference was found for teacher practice. The average observation score (on a 5-point scale ranging from 0 to 4) was 2.33 was for teachers at treatment group schools and 1.98 for teachers at control group schools, resulting in a statistically significant difference of 0.35 (effect size = 0.66, p = .018). This result indicates that after nearly two years of implementation, the intervention had statistically significant impacts on classroom practice of teachers in Hawai‘i.

Table 6.2 Estimated impact of Pacific CHILD on knowledge and practice of teachers in Hawai‘i

Outcome measure (corresponding exploratory question number)	Regression-adjusted means					Effect size	Number of observations
	Treatment schools	Control schools	Difference	Standard error	*p*-value		
Teacher knowledge: total score on knowledge assessment (exploratory question 3)	28.3	27.8	.43	.928	.640	.11	116
Teacher practice: average score on classroom observation (exploratory question 4)	2.33	1.98	.35*	.150	.018	.66	117

*Significant at the .05 level (two-tailed test).

Note: Regression-adjusted means were computed at the means of the covariates. Effect sizes were calculated by dividing the impact estimate by the standard deviation of the control group. The hierarchical linear modeling models corresponding to exploratory questions 3 and 4 were estimated using the restricted maximum likelihood method.

Source: Authors’ calculations based on data collected for this study.

[91] See appendix C for additional discussion of the selection of covariates included for the estimation results reported in table 6.2. Table C.3 of appendix C summarizes the covariates included in the subgroup analyses.

[92] Baseline school-average test scores were excluded from the teacher knowledge estimation, because the model did not converge when they were included.

Impact of Pacific CHILD on students in American Samoa and the Commonwealth of the Northern Mariana Islands

Additional exploratory analyses were conducted to identify patterns of potential impacts in American Samoa and the Commonwealth of the Northern Mariana Islands (CNMI). Like the exploratory subgroup analyses of Hawai‘i, these subgroup analyses were not intended to produce results that were directly comparable with impacts in other entities or with the overall pooled impact. The purpose of the subgroup analyses was to explore and understand the program effect for each entity.

The study was not explicitly designed to study subsamples for entities other than Hawai‘i. Based on preliminary analyses of the data, it was determined that the sample sizes did not allow the teacher impacts to be reliably estimated separately by entity (the estimated minimum detectable effect size for the teacher subsample was 1.58 in American Samoa, 0.92 in the CNMI, and 0.71 for the two entities combined).

Two exploratory questions about the impact on students were investigated:

- Exploratory question 5: Did grade 5 students at schools in American Samoa that were offered Pacific CHILD for two years perform differently on assessments of reading comprehension Pacific CHILD grade 5 students at schools in American Samoa that were not offered Pacific CHILD?
- Exploratory question 6: Did grade 5 students at schools in the CNMI that were offered Pacific CHILD for two years perform differently on assessments of reading comprehension from grade 5 students at schools in the CNMI that were not offered Pacific CHILD?

Each question was investigated independently based on the hierarchical linear modeling model, using the same outcome measure (SAT 10 reading comprehension scale scores). The estimation model for American Samoa included the same set of student- and school-level covariates used in the confirmatory analysis model (gender; special education status; baseline school-average test score, school size, student-to-teacher ratio, and percentage of students eligible for free or reduced-price lunch; and assignment block indicators). For the CNMI, the estimated model included a student-level covariate for race/ethnicity in addition to the covariates used in the confirmatory analysis model. For this reason, the results for the CNMI are not directly comparable to those for American Samoa.

Table 6.3 Estimated impact of Pacific CHILD on reading comprehension of grade 5 students in American Samoa

Outcome measure (corresponding exploratory question number)	Regression-adjusted means					Effect size	Number of observations
	Treatment schools	Control schools	Difference	Standard error	*p*-value		
Stanford 10 Achievement Test reading comprehension scale score (exploratory question 5)	595.6	598.7	–3.0	6.2	.629	–.15	185

Note: Regression-adjusted means were computed at the means of the covariates. Effect sizes were calculated by dividing the impact estimate by the standard deviation of the control group. The hierarchical linear modeling models were estimated using the restricted maximum likelihood method.

Source: Authors' calculations based on data collected for this study.

For the American Samoa subsample, the estimated impact is not statistically significant (table 6.3). For the CNMI subsample, the estimated impact on the reading comprehension scores is 10.64 points (effect size = 0.36), statistically significant at the 5 percent level (p = .025) (table 6.4).

Table 6.4 Estimated impact of Pacific CHILD on reading comprehension of grade 5 students in the Commonwealth of the Northern Mariana Islands

Outcome measure (corresponding exploratory question number)	Regression-adjusted means					Effect size	Number of observations
	Treatment schools	Control schools	Difference	Standard error	*p*-value		
Stanford 10 Achievement Test reading comprehension scale score (exploratory question 6)	635.5	624.9	10.6*	4.8	.025	.36	692

*Significant at the .05 level (two-tailed test).

Note: Regression-adjusted means were computed at the means of the covariates. Effect sizes were calculated by dividing the impact estimate by the standard deviation of the control group. The hierarchical linear modeling models were estimated using the restricted maximum likelihood method.

Source: Authors' calculations based on data collected for this study.

Impact of Pacific CHILD on subscales of teacher practice

The confirmatory analyses indicate that Pacific CHILD had a statistically significant impact on teacher practice (see chapter 5). This finding is based on the teacher practice measure constructed from data collected using a classroom observation instrument based on the Sheltered Instruction Observation Protocol (SIOP). The practice measure used in the confirmatory analysis was the overall total score on the observation instrument.

Given the statistically significant impacts on the overall total score, this section explores impacts on subscales of the teacher practice measure. This analysis investigated how the intervention

affected different content areas covered in the observation protocol. It examined which areas of teacher practice Pacific CHILD might have affected most.

The classroom observation protocol used in this study consisted of 38 items, including 30 based on the SIOP instrument and 8 developed for this study. All items were scored using a five-point Likert scale ranging from zero to four. The SIOP items cover three main dimensions identified by its developers: preparation, instruction, and lesson review and evaluation (Guarino et al. 2001). The eight items developed for this study were designed to assess the reading comprehension and instructional strategies emphasized by Pacific CHILD. The exploratory analysis investigated whether Pacific CHILD had impacts on each of the three dimensions of the SIOP and on the additional items emphasized by Pacific CHILD. It also estimated impact for subareas under the instruction dimension. The following questions were explored:

- Exploratory question 7: Did grade 4 and grade 5 teachers in schools that were offered Pacific CHILD for two years perform differently from grade 4 and grade 5 teachers at schools that were not offered Pacific CHILD, as measured by subscores in the preparation dimension of the teacher practice assessment?
- Exploratory question 8: Did grade 4 and grade 5 teachers at schools that were offered Pacific CHILD for two years perform differently from grade 4 and grade 5 teachers at schools that were not offered Pacific CHILD, as measured by subscores in the instruction dimension of the teacher practice assessment and by subscores in the subareas (comprehensible input, strategies, interaction, practice and application, and lesson delivery) of the instruction dimension of the teacher prtacic4e assessment.
- Exploratory question 9: Did grade 4 and grade 5 teachers at schools that were offered Pacific CHILD for two years perform differently from grade 4 and grade 5 teachers at schools that were not offered Pacific CHILD, as measured by subscores in the review and evaluation dimension of the teacher practice assessment?
- Exploratory question 10: Did grade 4 and grade 5 teachers at schools that were offered Pacific CHILD for two years perform differently from grade 4 and grade 5 teachers at schools that were not offered Pacific CHILD, as measured by subscores in the additional section of the teacher practice assessment on the reading and instructional strategies emphasized by Pacific CHILD?

Each question was examined independently, using the same analytical model and sample used in the confirmatory analyses.

The reliability of subscales measured by Cronbach's alpha varied (see appendix C for the alpha for each subscale). Four of the subscales examined had reliabilities of 0.70 or higher. These subscales include two of the three main dimensions of the original SIOP protocol (preparation and instruction) and two of the subarea scores under the instruction dimension (strategies and lesson delivery).

Estimated impacts are statistically significant for all subscales with Cronbach's alpha equal to or greater than 0.70 (table 6.5). The average rating for the preparation dimension for the treatment group was 0.34 points higher for the treatment group than for the control group (effect size = 0.50, $p = 0.013$). Similarly, the average rating in the instruction dimension was 0.33 points higher (effect size = 0.51, $p = 0.016$) for the treatment group than for the control group. Within the instruction dimension, the estimated difference between the treatment and control group for

the subarea of strategies was 0.44 points (effect size = 0.47, $p = 0.037$) and the estimated difference for the subarea of lesson delivery was 0.31 points (effect size = 0.41, $p = 0.024$).

Table 6.5 Estimated impact of Pacific CHILD on teacher practice subscales with Cronbach's alpha equal to or greater than .70

Subscale component[a] (corresponding exploratory question number)	Cronbach's alpha	Regression-adjusted means of average rating of teacher practice observation subscale					Effect size
		Treatment schools	Control schools	Difference	Standard error	*p*-value	
Preparation dimension of instructional practice (exploratory question 7)	.72	2.38	2.04	.34*	.138	.013	.50
Instruction dimension of instructional practice and its subareas (exploratory question 8)							
Instruction dimension	.87	2.52	2.19	.33*	.135	.016	.51
Strategies subarea	.76	2.56	2.13	.44*	.210	.037	.47
Lesson delivery subarea	.70	2.88	2.58	.31*	.136	.024	.41

*Significant at the .05 level (two-tailed test)

Note: Number of observations = 189. Regression-adjusted means were computed at the means of the covariates. Effect sizes were calculated by dividing the impact estimate by the standard deviation of the control group. The hierarchical linear modeling models corresponding to exploratory questions 7 and 8 were estimated using the restricted maximum likelihood method.

a. Results are shown only for measures with Cronbach's alpha of .70 or higher. See appendix C for the results of other outcome measures with Cronbach's alpha of less than .70.

Source: Authors' calculations based on data collected for this study.

Impact of moderating factors on teacher impacts

The following exploratory questions examined whether Pacific CHILD had varying effects on teacher outcomes depending on teachers' experience and level of education. The exploratory analyses examined which teachers might have benefited most from Pacific CHILD. The following exploratory questions were explored:

- Exploratory question 11: Did impacts on the knowledge of grade 4 and grade 5 teachers at schools offered Pacific CHILD for two years vary by teachers' years of teaching experience?
- Exploratory question 12: Did impacts on the knowledge of grade 4 and grade 5 teachers at schools offered Pacific CHILD for two years vary by teachers' level of education?
- Exploratory question 13: Did impacts on the instructional practice of grade 4 and grade 5 teachers at schools offered Pacific CHILD for two years vary by teachers' years of teaching experience?

- Exploratory question 14: Did impacts on the instructional practice of grade 4 and grade 5 teachers at schools offered Pacific CHILD for two years vary by teachers' level of education?

Each question was investigated by estimating the same analytical model estimated for the confirmatory analysis but allowing the treatment effects to vary by years of experience or education level.[93] The effect of education level was assessed by examining whether teachers with advanced degrees (that is, degrees beyond a bachelor's degree) performed differently from teachers without advanced degrees.

There were concerns that, because of correlations between the entity and the level of moderating factors, the estimation of effects of moderating factors would detect mainly between-entity variation in teacher characteristics rather than the effects of the variation in individual teacher characteristics (for example, no teachers reported having completed less than a bachelor's degree in Hawai'i or the CNMI). Such concerns were addressed to the extent possible in the estimation based on the pooled data.[94] Additionally, the effects of moderating factors on program impacts were also explored by limiting the sample to one entity, Hawai'i (see table C.7 in appendix C for the results based on the Hawai'i sample).

The estimation results for the effects of moderating factors are shown in table 6.6. Rather than presenting the mean difference between treatment and control groups, it displays the estimated effects of an additional year of experience or an advanced degree on the impacts of Pacific CHILD on each teacher outcome measure. The results indicate that neither an additional year of teaching nor an advanced degree moderated the impacts of the intervention on teacher practice.

The results also indicate that having an advanced degree did not moderate the impact of the intervention on teacher knowledge. However, the impact on the teacher knowledge assessment score is statistically higher for more experienced teachers. The estimation results suggest that an additional year of experience was associated with a difference in impact of 0.21 points (p = .023).

Additional investigation indicated that the statistically significant moderating effect of experience came largely from teachers with 10 or more years of experience; an additional year of teaching did not have a significant effect on the impact of Pacific CHILD among less experienced teachers (see appendix C). Readers are cautioned against drawing conclusions about the precise point estimate for the potential moderating effects based on the findings reported here, as the study was not designed to allow a rigorous analysis of such effects. The findings indicate a positive association between teachers' experience and the effectiveness of the intervention on their knowledge. Additional research is needed to assess the observed association and investigate possible reasons for differential impacts of Pacific CHILD by teacher experience.

[93] The estimated model included the moderating factors (teaching experience and education) as covariates as well as the interaction terms between the assignment condition indicator and the moderating factors. The effects of teaching experience and education on the slope of the impact were regarded as fixed rather than treating treated as random.

[94] Fixed entity-specific slopes effects were ruled out in preliminary specification tests; the estimation models included fixed entity intercepts (see appendix B).

Table 6.6 Estimated impact of teacher experience and education on impact of Pacific CHILD on teacher outcomes

Outcome measure/ moderating factor (corresponding exploratory question number)	Marginal effects on impact	Standard error	*p*-value	Number of observations
Total score on knowledge assessment/one additional year of experience (exploratory question 11)	.21*	.093	.023	190
Total score on knowledge assessment/holding advanced degree/ (exploratory question 12)	–.40	1.496	.787	190
Average score on classroom observation/one additional year of experience (exploratory question 13)	.00	.010	.912	189
Average score on classroom observation/holding advanced degree (exploratory question 14)	.04	.163	.793	189

*Significant at the .05 level (two-tailed test).

Note: Results are based on pooled data. Advanced degree refers to degree beyond a bachelor's degree. Regression-adjusted means were computed at the means of the covariates. The hierarchical linear modeling models corresponding to exploratory questions 11–14 were estimated using the restricted maximum likelihood method. The marginal effect of 0.21 on program impacts for teacher knowledge, for example, indicates that the program impact on average-profile teachers with five years of teaching experience was 1.4 points (effect size = 0.25); the impact on average-profile teachers with six years of experience was 1.6 points (effect size = 0.29). For each outcome measure, the moderating effects of years of experience and education were estimated jointly in a single estimation model.

Source: Authors' calculations based on data collected for this study.

Chapter 7: Summary of study design, confirmatory findings, and study limitations

This chapter begins by briefly reviewing the study design and implementation of Pacific CHILD. It then summarizes the main confirmatory findings and identifies the study's limitations.

Review of study design and implementation of Pacific CHILD

Fifty-one schools were recruited to participate in the study. Six schools were removed from impact analyses because of teacher reassignment that potentially compromised the integrity of the experimental design. The sample for confirmatory analyses thus consisted of 45 schools (23 randomly assigned to the treatment group, which was offered the two-year Pacific CHILD professional development program, and 22 schools randomly assigned to the control group). The student impact sample consisted of 3,078 grade 5 students (1,587 in treatment group schools and 1,491 in control group schools). The teacher impact sample consisted of 236 grade 4 and grade 5 teachers (118 in treatment group schools and 118 in control group schools). These samples included students and teachers who could potentially have been exposed to the intervention for the full two years. The study did not define the student analysis sample to be a cohort taught specifically by the teacher analysis sample.

Pacific CHILD was offered to grade 4 and grade 5 teachers who taught English language arts at the treatment group schools during the 2007/08–2008/09 school years in American Samoa and the Commonwealth of the Northern Mariana Islands (CNMI) and during the 2008/09–2009/10 school years in Hawai‘i. On average, the 118 treatment group teachers were exposed to 15 days of the intervention, which was designed to provide 42 days of professional development per teacher intended by the program's developers. Over the two years of the intervention, 68 of the 118 teachers at treatment group schools participated in the program.

Pacific CHILD was designed to be adopted and implemented across the diverse Pacific region. Program impacts were thus evaluated based on samples of students and teachers pooled over the three study entities. To pool the data from the entities, the study used a weighted-average approach for the student impact analysis and a combined sample approach for the teacher impact analysis.

In accordance with Institute of Education Sciences guidelines, the primary focus of the study was the program's effectiveness with respect to the student outcome. Student achievement in reading comprehension was measured with national, norm-referenced tests. Teacher knowledge was measured with an assessment designed for this study; teacher practice was measured by classroom observation, adapted from an existing tool. The impact estimation used a hierarchical linear model to estimate the impacts of Pacific CHILD on the student and teacher outcomes. The study specified two-level models in which individuals (students and teachers) were nested within schools to account for the effects of clustering within each school.

Effect of Pacific CHILD on student achievement in reading comprehension

The primary impact analysis examined whether Pacific CHILD affected student achievement in reading comprehension. It investigated the following question:

- Did grade 5 students at schools that were offered Pacific CHILD for two years perform differently on assessments of reading comprehension from grade 5 students at schools that were not offered Pacific CHILD?

The difference in reading comprehension scores between the treatment and control group students was found statistically significant. The estimated average score was 634.3 for students at treatment schools and 629.0 for students at control schools (effect size = 0.244, $p = .017$). Sensitivity analyses yielded consistent estimation results. The study thus finds a statistically significant impact on achievement in reading comprehension of grade 5 students at schools offered Pacific CHILD.

Effect of Pacific CHILD on teacher knowledge and practice

The secondary impact analysis examined whether Pacific CHILD had impacts on teacher knowledge or practice, the immediate targets of the professional development program. It investigated the following question:

- Did grade 4 and grade 5 teachers at schools that were offered Pacific CHILD for two years perform differently from teachers at schools that were not offered Pacific CHILD on either an assessment of their knowledge of theories and strategies related to effective reading instruction (including English language learner-focused theories and strategies) or an assessment of their instructional practices for enhancing student reading comprehension (including English language learner-focused practices)?

After adjustment for multiple testing, the differences between the treatment and control group teachers on both teacher outcome measures were found statistically significant. For teacher knowledge, the estimated average total score on a 40-point test was 27.0 for teachers at treatment group schools and 25.0 for teachers at control group schools. This difference of 1.96 points on the teacher knowledge assessment was found statistically significant (effect size = 0.35, adjusted $p = .023$). For teacher practice, the average observation score (on a 5-point Likert scale ranging from 0 to 4) was 2.20 for teachers at treatment group schools and 1.85 for teachers at control group schools. This difference of 0.36 on the classroom observation rating was statistically significant (effect size = 0.64, adjusted $p = .006$). Sensitivity analyses yielded estimation results consistent with the benchmark results. Regarding the secondary research question, the study thus finds statistically significant impacts on both knowledge and practice of grade 4 and grade 5 teachers at schools offered Pacific CHILD.

Study limitations and caveats

The study has limitations that should be considered when reviewing the results. They include the limited generalizability of findings, the composition of the American Samoa sample, the validity of outcome measures, sample equivalence, and sample attrition.

Limited ability to generalize results

The three entities studied—American Samoa, the CNMI, and Hawai‘i —were purposively selected based on the availability of student outcome data and the administrative support for this study. Given the intentional selection process, this study's findings are not generalizable to the broader Pacific Region. Furthermore, because the study schools within each entity were also a convenience sample, the findings are not representative of the entities themselves. The loss of generalizability means that formal inferences about the effectiveness of Pacific CHILD beyond the study schools in the three entities studied cannot be made. The impact findings cannot be used to extrapolate the program effects to other schools or entities in the past or the future.

The study was designed to test the intent-to-treat effects of Pacific CHILD on students and teachers at schools at which the program was offered. Although the underlying theoretical model of Pacific CHILD is based on assumptions about individual-level responses to the intervention, the study did not directly address the question of whether Pacific CHILD had an impact on individuals who were offered the intervention. Instead, it explored the question of whether Pacific CHILD had impacts on students and teachers at schools that were offered the intervention. For this reason, findings from this study are not intended to support conclusions about the intent-to-treat effects on individuals.

The possible lack of representativeness of the analysis samples also affects the ability to generalize outcomes to the study schools. All eligible teachers and students within each study school were targeted for inclusion in the impact samples in this study—that is, the samples of teachers and students were intended to represent the target populations at their schools. However, because of missing data caused by nonconsenters (teachers who explicitly declined to participate in the study) and nonrespondents (students who did not complete the assessments and teachers who failed to complete surveys or be observed), the actual analysis samples used in impact estimation may not have been fully representative of the target populations. This potential bias in the estimates could compromise the ability to generalize the results to the target populations in the study schools (see also the discussion below in the "section on sample attrition").

Composition of the American Samoa sample

Concerns about schools suspected of compromising the integrity of the experimental design led to the removal of six schools from the original sample recruited for this study, including larger schools in American Samoa. As a result, the schools American Samoa included in the impact sample were on average less than half the size of schools in the other entities. The study sample in American Samoa thus did not represent the full range of school types targeted by the original sample design. In interpreting the findings, readers should be aware of the smaller school size in American Samoa.

Validity of outcome measures

Although the measures used in this research are deemed valid and reliable for the target outcomes, they are not definitive measures for gauging the underlying constructs that Pacific CHILD was designed to affect.

For teacher knowledge and classroom practice, the study developed and adapted instruments based on a review of existing tools. Use of instruments developed or adapted for the study may raise concerns about overalignment with the intervention and the validity of the inferences based on data collected using the instruments. As discussed in chapter 3 and appendix D, steps, including reviews by external experts and pilot testing, were undertaken to address such concerns and ensure that the instruments developed would support valid inferences about impacts on teachers. Despite these steps, questions about overalignment cannot be completely addressed within the scope of this study. In addition, for the teacher practice measure, bias may arise due to the observers' knowledge of the assignment condition of schools. This could also potentially limit the study ability to support valid inferences about impacts.

For the student outcome measure, the instrument was selected from a short list of assessments that were already in use in the entities (the Stanford 10 Achievement Test [SAT 10] in American Samoa and the CNMI and the TerraNova in Hawai'i). Both assessments are nationally normed, standardized assessments whose psychometric properties, including validity and reliability, have been tested by their developers. These tests are commonly accepted and used by states and school districts across the country to assess students' academic performance; their content, developed in consultation with content experts, is considered grade appropriate and universally relevant (CTB/McGraw-Hill 2003; Pearson 2004). Nevertheless, concerns remain that these national norm-referenced tests may not be valid measures for assessing the effects of reading-focused interventions such as Pacific CHILD, especially in specific regional cultural contexts and for English language learner students. The use of these standardized tests in the study was justified on the grounds that they are widely used measures of reading achievement and regarded as policy-relevant tools.

Sample equivalence

Maintaining the sample equivalence in expectation across the conditions, except for the intervention, is critical for the experimental design of the study. To construct individual-level impact samples, this study targeted all study-eligible individuals who were enrolled or teaching at the study schools toward the end of the two-year intervention. These targeted students and teachers were considered comparable as groups across the conditions; thus, the cross-condition equivalence in the final analysis samples hinges on how outcome data were collected from these targeted individuals.

Systematic differences between the impact analysis samples across the treatment and control conditions, could lead to potential bias in the impact estimates. This was of a particular concern for teachers in this study. Given that the teacher impact sample was a cohort defined toward the end of the two-year intervention, factors such as teachers' knowledge of their schools' assignment status and their own exposure to the intervention to date—i.e., factors that differed across the conditions—could have influenced their participation in the outcome data collection efforts. If treatment group teachers who selected to participate in the outcome data collection efforts differed both in observed and unobserved ways from control group teachers who selected

to participate, the assumption of sample equivalence in expectation across the conditions could be compromised, leading to potential bias in the study results. As discussed below, one indicator for assessing possible compromise of cross-group equivalence is sample attrition. The observed level and patterns of attrition warrant concerns for potential bias in the teacher impact estimates.

Sample attrition

For the individual-level samples, determined at the time of the follow-up data collection, sample attrition was assessed in terms of the rates of data collection completion. (In case of the teacher impact sample, attrition accounts for both those who did not consent to participate in the study as well as those who consented to participate in the study but did not provide data.) Minimizing overall attrition within the individual samples (that is, maximizing the number of responses in data collection and minimizing missing data) and minimizing systematic difference in attrition across the treatment and control conditions increases the analytic sample size and decreases the risk of potential bias in impact estimates. The extent to which data collection was completed for the target sample and the extent to which data were collected consistently across the conditions are therefore important indicators in assessing the reliability of the study's findings.

One approach to evaluating the risk of bias in impact estimates caused by sample attrition is to apply the attrition bias model described in the *What Works Clearinghouse Procedures and Handbook* (U.S. Department of Education 2008) to the observed overall attrition and differential attrition across conditions. For the student sample, missing outcome data were minimal and balanced across conditions (the estimated overall attrition rate was 1 percent, and the differential attrition rate was 1 percent). According to the What Works Clearinghouse guidelines, the combination of the overall and differential attrition rates for student data in this study resulted in an acceptable level of bias (below the bias threshold of 0.05 standard deviation of the outcome). For the teacher sample, the overall attrition rate was 16 percent for both teacher outcomes; the differential attrition rate was 5.9 percent for the knowledge assessment and 5.0 percent for the practice observations. This difference, for example, could be due to their knowledge of whether or not the school was offered the intervention at the time when they were invited to participate in the study. Based on the attrition bias model in the What Works Clearinghouse guidelines, the combination of the overall and differential attrition rates for the teacher data resulted in an acceptable level of bias for the practice observations and a potentially acceptable level of bias for the knowledge assessment. Given the level of bias between acceptable and potentially acceptable for the teacher knowledge assessment, readers are cautioned to consider potential attrition bias in evaluating the teacher impact analysis results.

Movement of individuals in and out of the study schools during the intervention period also affected the composition of the impact analysis samples, which were defined as students and teachers at the study schools toward the end of the two-year intervention. If differential turnover leads to comparison groups that are no longer equivalent in expectation, impact estimates could be biased, limiting the study's ability to produce reliable findings. On the other hand, any intervention-induced movement of individuals could be considered part of the treatment condition, reflected in the intent-to-treat effects of the school-level intervention.

Appendix A: Statistical power analysis

This appendix presents the statistical power analyses that were conducted during the design phase of the study. It describes the a priori analyses conducted before schools were recruited, in order to establish the expected power of the study and determine the target sample size. The appendix also presents the statistical analyses conducted after two blocks of the original sample were dropped from the impact analyses. This analysis was conducted to check and confirm that the reduced sample had sufficient statistical power to detect the impacts reliably. The Optimal Design software package developed by Stephen Raudenbush and others was used to conduct the power analyses (Spybrook et al. 2009).

The study adopted a cluster random assignment design, with the school as the unit of randomization and individuals within the school as the unit of analysis. The study was designed to offset the effects of clustering on statistical power through statistical adjustments. Previous studies have shown that controlling for cluster-level variation in estimating impacts can mitigate these effects in studies that use a cluster random assignment (Schochet 2005; Bloom et al. 2005; Bloom, Bos, and Lee 1999). Bloom, Bos, and Lee (1999) find that clustering effects can be reduced dramatically by including baseline covariates in regression estimations of impacts. In their study of a large urban school district, they find that the use of school-level reading test scores from a previous cohort of students reduced clustering effects. Inclusion of the baseline cluster-level aggregate test score reduced the effective intraclass correlation coefficient from .20 to .05. Based on these studies, this study was designed to compute regression-adjusted estimates for impacts using individual- and school-level measures as covariates. The a priori power analyses for this study assumed that the effective intraclass correlation coefficient would be reduced to .05 by applying an appropriate regression adjustment.

The assumptions made in computing estimated the minimum detectable effect size were as follows:

- Statistical power of 0.80.
- 5 percent level of significance ($\alpha = 0.05$) for a two-tailed test.
- Intraclass correlation of .05.
- Explanatory power (R^2) of individual-level covariates of .5 for student outcomes and 0 for teacher outcomes

The goal in making decisions about the target sample size (and statistical power) was to attain minimum detectable effect sizes that were sufficiently small to allow reliable detection of impacts but large enough to be meaningful from a policy perspective. The analyses indicated that 50 schools were needed to yield a minimum detectable effect size of 0.15 for student outcomes and 0.40 for teacher outcomes for analyses of the full sample (table A.1). These minimum detectable effect sizes were determined to be sufficiently small for this study. Based on these analyses, the recruitment goal was set at 50 or more schools.

Table A.1 Power analysis based on original assumptions about sample size for measuring impact of Pacific CHILD in study schools

Outcome	Number of schools	Students or teachers per school	Minimum detectable effect size
Student outcomes (R^2 = 0.5)			
Full sample	50	125	.15
Subsample of 25 Hawai‘i schools	25	125	.21
Teacher outcomes (R^2 = 0)			
Full sample	50	5	.40
Subsample of 25 Hawai‘i schools	25	5	.57

Note: Power = 0.8, significance level = 0.05, intraclass correlation coefficient = 0.5.

Source: Authors' calculations using Optimal Design software.

During the first year of implementation, the study team learned that several treatment group teachers had been replaced to accommodate the implementation of Pacific CHILD. Because such reassignment could compromise the integrity of the experimental design, the affected schools were dropped, along with the other schools in the same assignment blocks, from impact analyses. At the time this decision was made, recruitment in Hawai‘i was still underway (see chapter 2). In order to determine whether more schools needed to be recruited, additional power analyses were conducted based on revised assumptions that took into account the information on the schools that had been recruited through that point. The analysis indicated that one additional school needed to be recruited in Hawai‘i to ensure that the sample size had sufficient power to reliably estimate impacts (table A.2). The reduced sample size led to a loss of power, with a minimum detectable effect size of 0.16 for student outcomes and 0.46 for teacher outcomes, which were still deemed to be sufficiently small to produce meaningful findings.

Table A.2 Power analysis based on revised assumptions about sample size for measuring impact of Pacific CHILD in study schools

Outcome	Number of schools	Students or teachers per school	Minimum detectable effect size
Student outcomes (full sample, $R^2 = 0.5$)	45	110	.16
Teacher outcomes (full sample, $R^2 = 0$)	45	4.5	.46

Note: Power = 0.8, significance level = 0.05, intraclass correlation coefficient = .05.

Source: Authors' calculations using Optimal Design software.

Appendix B: Technical notes on impact analyses

This appendix provides technical details on the confirmatory impact analyses based on the randomized control trial of Pacific CHILD. It describes the statistical models, covariates included in the estimation, approaches to pooling data across the entities, treatment of missing data and other data issues, and sensitivity analyses.

Analytic approach

This section presents the statistical model and data used to conduct the impact analyses reported in chapter 5.

Tests for impacts

The purpose of the impact analyses was to draw statistical inferences on the effects of a randomized offer of Pacific CHILD on student and teacher outcomes. To answer the primary research question, regarding impact on students, the study tested the following null hypothesis:

$$H_0: \gamma_{1(\text{STUDENT})} = 0$$

where $\gamma_{1(\text{STUDENT})}$ represents the estimated intent-to-treat effect on student reading comprehension. If the null hypothesis was rejected by a two-tailed test at the 5 percent significance level, the study would conclude that the outcome was different for students at schools that were randomized to receive an offer of Pacific CHILD and students at schools that were not offered Pacific CHILD.

To answer the secondary research question, regarding teacher outcomes, the study test the following global null hypothesis:

$$H_0 = \{ [H_{0k}: \gamma_{1(\text{KNOWLEDGE})}) = 0] \text{ and } [H_{0p}: \gamma_{1(\text{PRACTICE})} = 0] \}$$

where $\gamma_{1(\text{KNOWLEDGE})}$ represents the estimated intent-to-treat effect on teacher knowledge and $\gamma_{1(\text{PRACTICE})}$ represents the estimated intent-to-treat effect on teacher practice. The secondary question examines the effectiveness of the intervention across the two outcome domains. The null hypothesis H_0 is rejected if either H_{0k} or H_{0p} is rejected based on independently conducted significance tests, correcting for multiple comparisons. The method proposed by Benjamini and Hochberg (1995) was applied to control for the false discovery rate.[95] If either domain-specific null hypothesis was rejected based on the *p*-value adjusted for multiple comparisons by a two-tail

[95] In general, the Benjamini-Hochberg procedure for adjusting for *M* (the number of outcomes) comparisons entails first conducting *M* tests separately at the significance level α and ranking *p*-values from these tests from smallest to largest, $p_1 \leq \ldots p_m \ldots \leq p_M$, where *m* represents the order of the test based on the *p*-value. All null hypotheses for $m = 1, 2, \ldots k$ are rejected (that is, impact estimates with which these tests are associated are deemed statistically significant) where *k* is the maximum *m* such that $p_m \cdot M/m \leq \alpha$. All null hypotheses for $m = k+1, \ldots M$ are not rejected (that is, impact estimates with which these tests are associated are deemed not statistically significant). The adjusted *p*-values used in the stepwise procedure (computed by multiplying the *p*-values by *M/m*) are provided in reporting the application of the Benjamini-Hochberg procedure. In this study $M = 2$ and $\alpha = 0.05$. Therefore, the Benjamini-Hochberg procedure adjusted the two *p*-values by multiplying the smaller value by 2 and the larger value by 1. If either of the adjusted *p*-values was .05 or less, the null hypothesis that there were no impacts on any of the teacher outcomes was rejected.

test at the 5 percent significance level, the study would conclude that the intervention had impacts on at least one of the teacher outcomes.

Basic statistical model

The estimation of the program effects, γ_1, was based on hierarchical linear modeling.[96] The model was specified as a two-level random-intercept model, in which the first level (student or teacher level) was nested in the second level (school level). In particular, for individual (student or teacher) i and school j, for $i = 1\ldots N$ and $j = 1\ldots K$, the hierarchical model was specified as the following system of equations:

$$\begin{cases} Y_{ij} = \alpha_j + \sum_{q=1}^{Q} \beta_q X_{qij} + \varepsilon_{ij} & \text{Level 1 (individual level)} \quad \text{(Equation B.1)} \\ \alpha_j = \gamma_0 + \gamma_1 (STATUS)_j + \sum_{s=2}^{S} \gamma_s W_{sj} + u_j & \text{Level 2 (school level)} \quad \text{(Equation B.2)} \end{cases}$$

where Y_{ij} is an outcome measure for student or teacher i in school j; X_{qij} is the qth individual-level covariate for observed baseline characteristics, for $q = 1\ldots$ Q; *STATUS* is a dummy variable indicating whether school j was randomly assigned to receive Pacific CHILD (*STATUS* = 1) or not (*STATUS* = 0); W_{sj} is the sth school-level covariate, for $s = 2\ldots S$. In addition to these observed variables, the model specifies parameters (α, β, γ) to be estimated, where α_j is the school-specific intercept, representing the adjusted mean outcome for school j; γ_0 represents the adjusted mean outcome across control group schools (when *STATUS* = 0); γ_1 is the impact estimator, representing the regression-adjusted mean difference in outcomes between treatment and control group schools; and β_q and γ_s are estimators for marginal effects of individual- and school-level covariates. The effects of the individual-level covariates, β_q for $q = 1\ldots$ Q, were constrained to be fixed across the school level (Level 2). The model assumes two random error terms: ε_{ij} is the residual term specific to student i in school j; u_j is the residual specific to the jth school. The error terms ε_{ij} and u_j are assumed to be independently and normally distributed, each with mean 0 and constant variance (σ_ε^2 and σ_u^2), such that $\varepsilon_{ij} \mid u_j \sim N(0, \sigma_\varepsilon^2)$ and $u_j \sim N(0, \sigma_u^2)$.

Substituting equation B.2 into equation B.1 allows the system of equations to be rewritten in reduced-form format:

$$Y_{ij} = \gamma_0 + \gamma_1 (STATUS)_j + \sum_{q=1}^{Q} \beta_q X_{qij} + \sum_{s=2}^{S} \gamma_s W_{sj} + \delta_{ij} \qquad \text{(Equation B.3)}$$

where $\delta_{ij} = \varepsilon_{ij} + u_j$ represents the unobserved residual term. Because of the school-level random effect u_j, the error term δij in equation B.3 is not homoscedastic. Consequently, the standard ordinary least squares estimator would be unbiased but no longer efficient, and the standard errors of the ordinary least squares estimators and statistical inferences based on them would not be correct. As discussed below, the model is therefore estimated using alternative estimation methods.

[96] For additional discussions of specifications of the hierarchical linear model, see Raudenbush and Bryk (2002).

One extension of the model that was considered was to allow the coefficient on the impact estimator to vary across schools (that is, to assume a random slope on the *STATUS* variable).[97] This random coefficient model was rejected based on preliminary tests of the variance components (see table B.1). For the outcome analyses reported, therefore, it was assumed that the treatment effect did not vary by school; the results from the models are presented without random coefficients.

Table B.1 Preliminary specification tests: *p*-values from likelihood ratio test

Fully unrestricted model:

$$Y_{ij} = \gamma_0 + \lambda_{AS} AS + \lambda_{NM} CNMI + \gamma_1 (STATUS)_j + \varphi_{AS} AS(STATUS)_j + \varphi_{NM} CNMI(STATUS)_j + u_j + \mu_j (STATUS)_j + \varepsilon_{ij}$$

	Restricted model (M_R)	Unrestricted model (M_U)	*p*-value for likelihood ratio test by outcome measure type		
			Teacher knowledge	Teacher practice	Student achievement in reading
1.	Ordinary least squares: $\lambda = \varphi = \mu = u = 0$	Base model: $\lambda = \varphi = \mu = 0$	< .000	< .000	< .000
2.	Base model: $\lambda = \varphi = \mu = 0$	Base plus random slope: $\lambda = \varphi = 0$	1.000	.983	.813
3.	Base model: $\lambda = \varphi = \mu = 0$	Base plus fixed entity effects (benchmark): $\varphi = \mu = 0$	< .000	< 0.000	< .000
4.	Base plus fixed entity effects (benchmark): $\varphi = \mu = 0$	Benchmark + fixed entity–specific slope: $\mu = 0$	.540	.837	.863

Note: Test statistics were based on maximum likelihood estimation results. Likelihood ratio tests were used to test the difference between a restricted model and an unrestricted model (a restricted model is nested in an unrestricted model).

Source: Authors' calculations based on data collected for this study.

Pooling data across entities

The basic model was extended to incorporate the multisite evaluation design, in which the study schools were sampled from three entities across the Pacific region. These entities were selected deliberately rather than randomly, taking into consideration factors such as the availability of data and the level of support from entity educational agencies. Because of the purposive nature

[97] This random coefficient model can be expressed as follows:

$$Y_{ij} = (\gamma_0 + u_j) + (\gamma_1 + \mu_j)(STATUS)_j + \sum_{q=1}^{Q} \beta_q X_{qij} + \sum_{s=2}^{S} \gamma_s W_{sj} + \varepsilon_{ij}$$

$$= \gamma_0 + \gamma_1 (STATUS)_j + \sum_{q=1}^{Q} \beta_q X_{qij} + \sum_{s=2}^{S} \gamma_s W_{sj} + (\delta_{ij} + \mu_j (STATUS)_j)$$

where μ_j is the (unobserved) random school effect on the impact estimator, with the assumption that given $STATUS_j$, u_j and μ_j have a bivariate normal distribution with zero means and a symmetric covariance matrix. This random coefficient model was tested against the basic model for each outcome measure. The null hypothesis that $\mathrm{Var}(\mu_j \mid STATUS_j) = \mathrm{Cov}(u_j, \mu_j \mid STATUS_j) = 0$ for each outcome measure was not rejected based on likelihood ratio tests.

of the selection of the three entities, the study was not designed to use the data from these entities to draw inferences about the region in general.

Pacific CHILD was designed to be adaptable to varying local contexts and to be effective across the entities in the region. The study was therefore designed to evaluate the effectiveness of the intervention, pooling the data from all the three entities. To pool the data over the three entities, the study considered two approaches. One approach was to estimate overall impacts based on a combined sample consisting of observations from all entities. Another approach was to estimate overall impacts as the weighted average of entity-specific impacts that were estimated separately within each entity. The weighted-average approach was adopted for the student data; the combined sample approach was used to pool the teacher data.

Teacher data

In the combined sample approach adopted for the teacher impact analyses, each individual record from all entities was given an equal weight in estimating the program effects. Potential entity-to-entity variation was accounted for by including fixed entity effects in the model. The fixed entity effects could be entered as intercepts as well as slopes of the impact variable. The study team concluded that the fixed entity-specific slopes (that is, the interaction terms between the entity indicators and the treatment indicator) did not contribute to the estimation of any of the three outcome measures once the fixed entity-specific intercepts were included. The basic model with the entity-specific fixed intercepts can be expressed as follows:

$$Y_{ij} = \lambda_{HI} HI + \lambda_{AS} AS + \lambda_{NM} CNMI + \gamma_1 (STATUS)_j + \sum_{q=1}^{Q} \beta_q X_{qij} + \sum_{s=2}^{S} \gamma_s W_{sj} + \delta_{ij} \quad \text{(Equation B.4)}$$

where *HI* is the dummy entity indicator for Hawai‘i; *AS* is the dummy entity indicator for American Samoa; *CNMI* is the dummy entity indicator for the Commonwealth of the Northern Mariana Islands; and λ_{AS}, λ_{NM}, and λ_{HI}, are the parameters measuring the fixed entity-specific effects for American Samoa, the CNMI, and Hawai‘i.

A summary of the results from preliminary specification tests is provided in table B.1. Based on these analyses, equation B.4 was determined as the benchmark model form for the teacher outcome analyses. In addition to school- and individual-level covariates, the benchmark model included the school-specific random intercept term and entity-specific fixed intercepts to account for between-entity variation in the outcome measures. A set of dummy variables was also included to indicate the assignment blocks used to stratify schools during random assignment. Because schools were blocked within each entity (see chapter 2), the inclusion of entity indicators and assignment blocks would result in perfect collinearity. In practice, therefore, a noncollinear subset of these indicators was included in the estimation.

Student data

For the student outcome, additional preliminary analyses were conducted to check variation across entities by estimating the effects separately.[98] The results suggested that the impact

[98] Although the likelihood ratio test indicated that fixed entity-specific slopes did not make additional contribution to the estimation of the student outcome measure once entity-specific intercepts were included in the model, the fixed entity-

estimates may have varied considerably by entity. Given this potential variation across entities, an alternative approach was used to pool the data in which the program effect was estimated as a weighted average of the three entity-specific impacts that were independently estimated using the same set of covariates. Specially, weights (w) were defined as follows:

$$w_k = \frac{1}{se_i^2} \qquad \text{(Equation B.5)}$$

where se_i is the standard error of the student effect estimate γ_{1k} in entity k. The weighted-average estimate, γ_1, and its variance, $v(\gamma_1)$, were calculated as:

$$\gamma_1 = \frac{\sum_k w_k \gamma_{1k}}{\sum_k w_k} \qquad \text{(Equation B.6)}$$

and

$$v(\gamma_1) = \frac{1}{\sum_k w_k} \qquad \text{(Equation B.7)}$$

This weighted-average approach gives more weight to more precise estimates and less weight to less precise estimates.[99] As in the teacher impact analyses, the overall effect estimate for the student outcome thus reflected the effectiveness of the program measured across the three entities; the averaged effect took into account the variation in the impact estimate across entities.[100] For comparison, the impacts on the student outcome were also assessed based on the combined sample approach (the results are reported below).

Data

For the student outcome measure, the study used the grade 5 reading comprehension scores on the Stanford 10 Achievement Test (SAT 10) (or SAT 10–equivalent) administered at study schools toward the end of the two-year intervention (see appendix E for additional information on the student outcome measure). Based on enrollment information collected separately, it was estimated that the test data were collected from 99 percent of students enrolled at the study schools at the time of data collection.[101] Listwise deletion was used to handle missing student test data, assuming that outcome data were assumed to be missing completely at random.

---

specific slope estimate for American Samoa was statistically significant; the fixed entity-specific slopes for the two other entities were not statistically significant.

[99] This approach is frequently used in meta-analysis to compute weights for combining effects across independent samples (Cooper, Hedges, and Valentine 2009).

[100] The weighted-average approach was not used for teachers, partly because effects for American Samoa and the CNMI could not be reliably estimated because of the smaller teacher sample sizes.

[101] The analytical sample was defined as a cross-sectional group of grade 5 students enrolled at the study schools at the time of testing.

Table B.2 Covariates used to estimate impact of Pacific CHILD in confirmatory analyses

Level/type of covariate	Variables
School (Level 2)	
Assignment condition	▪ Treatment group indicator (*STATUS*)
Assignment block and entity indicator	▪ Assignment block dummies and entity indicators (noncollinear subset of indicators)[a]
School performance	▪ Baseline year average scores on reading comprehension subtest of Stanford 10 Achievement Test (SAT 10)[b]
School characteristics	▪ School size (number of students in school) at baseline[c] ▪ Student-to-teacher ratio at baseline ▪ Percent free or reduced-price lunch-certified students at baseline
Teacher (Level 1)	
Demographics	▪ Gender (binary indicator for male) ▪ Primary language (binary indicator if primary language other than English) ▪ Race/ethnicity (binary indicator for White)
Experience/qualifications	▪ Total years of teaching at baseline ▪ Years at current school at baseline ▪ Grade taught (binary indicator for grade 4) ▪ Highest degree completed (binary indicators for completing less than a bachelor's degree and completing more than a bachelor's degree) ▪ Certification status
Student (Level 1)	
Demographics	▪ Gender (binary indicator for female) ▪ Special education status (binary indicator)

a. Because schools were blocked within entities, assignment indicators and entity indicators would result in perfect collinearity. Entity indicators were therefore included only in the teacher outcome analyses.

b. The baseline year school average was computed based on grade 5 scores for schools in Hawai'i and the CNMI and grade 4 scores for schools in American Samoa (where grade 5 scores were not available). For Hawai'i, TerraNova scores were used to estimate equivalent SAT 10 scores, using the methodology described in appendix E

c. The number of grades varied by school. The study examined the number of fourth and fifth graders per school, which was collected during the recruitment period, as a secondary source of the baseline school size. Estimations based on the fourth and fifth grade size yielded the same results as those based on the total school size.

Source: Authors.

Data for the teacher outcome measures were collected using the teacher knowledge assessment and teacher practice observations toward the end of the two-year intervention (see appendix D for additional information on the teacher outcome measures). Outcome data were collected from 84 percent of teachers who met the study criteria (see chapter 3). Listwise deletion was used to handle missing teacher outcome data (if any outcome variable was missing for an individual, the individual was removed from the analysis). These outcome data were assumed to be missing completely at random.

The estimation models included both school- and individual-level covariates, including baseline and background variables as well as indicators for a set of assignment block indicators (table

B.2; see chapter 3 for descriptions of the data sources for the school- and individual-level covariates).

Individual-level student and teacher background data were collected after random assignment. Some of these data, such as gender, race/ethnicity, special education status, and primary language, were time invariant and independent of the intervention; they were considered baseline measures, even though they were collected after random assignment. A recent study that examines the use of late pretests in randomized control trials (Schochet 2008) concludes that including postassignment data as pretest proxies in analyses is preferable to excluding them, even though such data are not considered independent of the intervention. Listwise deletion was used to handle missing covariates.[102] The dummy variable adjustment method was considered as an alternative approach to account for the missing covariates, as discussed below. Summary statistics for the outcome variables and covariates included in the analyses are shown in table B.3 (see tables 5.1 and 5.3 in chapter 5 for the summary statistics for the outcome variables for the control groups).

[102] The background data and outcome data collected during the second year of the intervention were used as the primary source of information for constructing covariates. If the data were missing for the second year but available for the first year of data collection, data from the first year data were used.

Table B.3 Summary of outcome variables and covariates used to estimate impact of Pacific CHILD in confirmatory analyses

Variable or covariate	Mean	Standard deviation	Minimum	Maximum	Number of observations
Outcome variables					
Stanford 10 Achievement Test reading comprehension scale score	633.8	37.97	552	734	3,052
Total score from knowledge assessment	25.93	5.38	8	35	197
Average score from classroom observation	2.01	0.57	0.63	3.36	198
School-level covariates					
School size	474.6	243.8	< 70	> 1,000	45
Student-to-teacher ratio (number of students per teacher)	15.9	3.2	< 7.0	> 21.0	45
Percentage of students receiving free or reduced-price lunch	0.69	0.29	< 0.13	> 0.99	45
Average baseline reading score, used in student estimation[a]	625.5	25.0	< 570	> 650	45
Average baseline reading score, used in teacher estimation[a]	619.9	22.2	< 570	> 640	45
Teacher-level covariates					
Male	0.21	0.40	0	1	200
Years of teaching (at baseline)	8.60	8.01	0	> 35	194
Years at current school (at baseline)	4.70	5.67	0	> 23	193
Completed less than bachelor's degree	0.05	0.22	0	1	198
Competed more than bachelor's degree	0.32	0.47	0	1	198
Has a credential	0.98	0.14	0	1	196
Primary language not English	0.12	0.33	0	1	197
Teaching grade 4	0.49	0.50	0	1	236
Asian	0.36	0.48	0	1	198
Pacific Islander	0.35	0.48	0	1	198
White	0.40	0.49	0	1	198
Student-level covariates					
Female	0.48	0.50	0	1	3,052
Special education student	0.11	0.31	0	1	3,052

a. Baseline student scores were available for grade 4 in American Samoa, grade 5 in the CNMI, and grades 4 and 5 in Hawai'i. For the student impact estimation for the sample of grade 5 students, baseline scores were constructed based on grade 4 data for American Samoa and grade 5 data for Hawai'i and the CNMI . For the teacher impact estimation, based on the sample of grade 4 and grade 5 teachers, baseline scores were constructed based on data for students in grade 4 in American Samoa, grade 5 in the CNMI, and grades 4 and 5 in Hawai'i.

Source: Authors' computation based on data collected for this study.

Impact estimation results

In the random intercept model described above, the effects of clustering of students and teachers at schools were explicitly specified, in the form of between-school heterogeneity represented by school-specific random intercepts (u_j). The cluster-adjusted covariance parameters to be estimated were then derived along with regression coefficients. Restricted maximum likelihood and (full) maximum likelihood methods were considered in estimating this model.[103] Both estimators are efficient and generate unbiased estimates for regression coefficients. The restricted maximum likelihood method was selected as the benchmark estimation method. Unlike the maximum likelihood method, it takes into account the loss of degrees of freedom that results from estimating the fixed-effects parameters in estimating the variance components (and provides an unbiased estimate for the variance components for balanced data). Consequently, the restricted maximum likelihood method yields more conservative (larger) estimates for standard errors for regression coefficients than the maximum likelihood method.[104] The study also estimated impact based on the maximum likelihood method as well as other alterative estimation methods to check whether the estimates were sensitive to the choice of the estimation methods. The study reports the impact estimates based on the restricted maximum likelihood method as the main findings of the study.

Unadjusted estimates of impacts

The impact estimates unadjusted for individual- and school-level covariates are shown in table B.4. These estimates accounted for clustering effects at the school level. Without controlling for baseline covariates, the differences between treatment and control groups in student reading comprehension and teacher knowledge scores are not statistically significant. The difference in teacher practice scores is statistically significant (difference = 0.31, effect size = 0.55, p = .006).

[103] The study team reviewed a number of documents on estimation methods used in clustered randomized controlled trials, including Schochet (2009) and West et al. (2007).

[104] The maximum likelihood methods were used to conduct preliminary specification analyses based on likelihood tests.

Table B.4 Unadjusted means of outcome measures in treatment and control schools

Outcome measure	Treatment schools	Control schools	Difference	Standard error	*p*-value	Effect size	Number of observations
Stanford 10 Achievement Test reading comprehension scale score							
All entities combined	628.4	629.4	–.97	5.346	.856	–.03	3,052
American Samoa	594.7	600.5	–5.84	4.046	.149	–.28	185
Commonwealth of Northern Mariana Islands	629.9	633.4	3.50	5.347	.513	.12	692
Hawai‘i	638.0	636.7	1.37	4.489	.771	.04	2,175
Total teacher knowledge assessment score	26.7	25.3	1.40	1.137	.218	.25	197
Average score from classroom observation of teachers	2.19	1.88	0.31**	0.112	.006	.55	198

**Significant at the .01 level (two-tailed test).

Note: Scores are based on reading comprehension assessment data from the Stanford 10 Achievement Test (SAT 10) for American Samoa and the CNMI and the TerraNova for Hawai‘i. TerraNova scores were converted to SAT 10–equivalent scores using the methodology described in appendix E. Unadjusted means were attained by estimating multilevel models that account for clustering of individuals within schools with the assignment indicator but no other covariates. Effect sizes were calculated by dividing the impact estimate by the standard deviation of the control group.

Source: Authors’ calculations based on data collected for this study.

Primary impact analyses of student outcomes based on the benchmark model

For the student impact estimation, the covariates accounted for part of the between-school variance in test scores, especially in the CNMI and Hawai‘i. The unconditional intraclass correlation estimated based on the variance-component model was .04 for American Samoa, .06 for the CNMI, and .07 for Hawai‘i. Once observable school and student characteristics were controlled for, the conditional intraclass correlation was reduced to .04 in the CNMI and .01 in Hawai‘i; it remained unchanged in American Samoa.

For each entity, the benchmark hierarchical linear modeling model with school- and teacher-level covariates was independently estimated. Using the weights described above, the study computed the overall impact estimate across the three entities as a weighted mean of the entity-specific estimates. The weighted-average results are reported in chapter 5 and in table B.5, along with the results from the estimation results by entity. These results show variation in impact estimates across entities, supporting the weighted-average approach for pooling data. The weighted-average estimate of the impacts are statistically significant at the 5 percent level.

For comparison, table B.5 also presents the estimates based on the combined sample of all students from the three entities, which included the fixed entity effects in the model. The estimation based on the combined sample finds smaller and statistically insignificant estimates; additional analyses find that the student impact estimates based on the combined sample are

sensitive to estimation methods. These contrasting results underscore the importance of selecting an appropriate method for pooling data across the three entities. This study adopted the commonly used weighted-average approach to take account of the observed variation across entities. The impact estimation results might have been different had the data been pooled differently or the entity-to-entity variation been addressed differently.

Table B.5 Estimated impacts of Pacific CHILD on grade 5 Stanford 10 Achievement Test reading comprehension scale scores

Estimate	Treatment schools	Control schools	Difference	Standard error	*p*-value	Effect size	Number of observations
Estimate of overall impact							
Weighted mean of three entity estimates (benchmark model)	634.3	629.0	5.3*	2.19	.017	.244	3
Entity and combined sample estimates							
American Samoa	595.6	598.7	–3.0	6.24	.629	–.146	185
Commonwealth of the Northern Mariana Islands	636.1	624.6	11.5*	5.23	.027	.392	692
Hawai‘i	640.6	635.5	5.1	2.62	.050	.133	2,175
Combined sample	635.2	632.6	2.5	2.18	.247	.068	3,052

*Significant at the .05 level (two-tailed test).

Note: Scores are based on reading comprehension assessment data from the Stanford 10 Achievement Test (SAT 10) for American Samoa and the CNMI and the TerraNova for Hawai‘i. TerraNova scores were converted to SAT 10–equivalent scores using the methodology described in appendix E. Results are based on restricted maximum likelihood estimation. For each entity, regression-adjusted means were computed at the means of the covariates, and effect sizes were calculated by dividing the impact estimate by the standard deviation of the control group. Each entity estimation included the following covariates: blocking variables, school-level baseline reading comprehension scale score, school size, student-to-teacher ratio, percentage of students eligible for free or reduced-price lunch, student gender, and student special education status. Overall impact and effect size were computed as weighted means of single-entity estimates.

Source: Authors’ analysis based on student records collected for this study.

Secondary impact analyses of teacher outcomes based on the benchmark model

The study finds that the impacts Pacific CHILD on both teacher outcomes are statistically significant. For the teacher outcome measures, unconditional intraclass correlation was estimated based on the variance-component model without any teacher- and school-level covariates. The unconditional intraclass correlation was .32 for the teacher knowledge assessment and .30 for teacher practice. Once observable school and student characteristics were controlled for in the estimation, the conditional intraclass correlation based on the benchmark model for teacher knowledge was reduced to .10; the conditional intraclass correlation for teacher practice remained at .30. For the benchmark specification, the covariates thus did not do a good job of explaining the between-school variance for the teacher practice outcome.

The results from the fully specified model for teacher outcomes are reported in chapter 5 and summarized in table B.7. For the benchmark model presented in these tables, the Benjamini-Hochberg procedure was used to adjust for multiple comparisons. The study team also considered the Bonferroni procedure, an alternative adjustment to control for the familywise error rate. For this adjustment, the confidence bound (in this case, 0.05) was divided by the number of tests compared (two) and applied as the significance for each test. The null hypothesis that there was an impact in neither teacher outcome domain was rejected based on the Bonferroni method as well as the Benjamini-Hochberg method.

Sensitivity analyses

Several sensitivity analyses were conducted to assess whether the results from the benchmark analytic model remained robust across a range of other methodological choices.

Alternative estimation methods

For the benchmark random intercept model, the study used the restricted maximum likelihood method to estimate the coefficients and covariance parameters derived from the hierarchical linear model. The random intercept model was also fit using the maximum likelihood method (which provides downward-biased estimates for σ_u^2) and feasible generalized least squares based on the Swamy and Arora ANOVA method, which provides an unbiased but not efficient estimator for σ_u^2. Two alternative approaches that do not explicitly estimate the within-school covariance structure were also considered: generalized estimating equations and ordinary least squares with robust (Huber-White) standard errors.[105] As the primary goal of the study was to estimate the regression coefficient on the treatment indicator, not the random effects variance component, these estimators were regarded as reasonable alternative approaches for checking the robustness of the impact estimation results.[106]

Alternative methods for treating missing observations in covariates

Listwise deletion was used to address missing covariates for the benchmark models for teacher outcomes. As part of the sensitivity analyses, the benchmark model was refit using the dummy variable adjustment methods for teacher outcomes.

Alternative specifications of outcome measures

For the benchmark analysis of student achievement, the equipercentile linking method was used to pool SAT 10 scores from American Samoa and the CNMI and TerraNova scores from Hawai‘i (see appendix E). As an alternatively scaled measure, *z*-scores were used to estimate the

[105] Generalized estimating equation parameters are estimated by an iterative optimization process, with the working covariance as a function of the working correlation matrix (of the dependent variable). The form of this working correlation matrix was assumed to be exchangeable. The covariance parameters are treated as nuisance variables in the iterative process. Estimates for the covariance based on a generalized estimating equation model are consistent, assuming the correlation matrix is correctly specified.

[106] See Schochet (2009) for a discussion of various estimation methods used in clustered randomized controlled trials.

model. Test scores were standardized using the mean and standard deviation of the scores within each entity to construct these scores.

For the benchmark analysis of teacher knowledge, the total score from the teacher knowledge assessment was used. To check the robustness of the benchmark findings, the study constructed an alternative outcome based on total score, excluding items that had low item discrimination and low item facility scores and did not contribute to the overall reliability of the outcome measure. An alternative outcome measure based on a two-parameter item response theory model was also used.

For the benchmark analysis of teacher practice, the average Likert scale score was used from the classroom observation protocol, adjusted for unrated items. To check that the results were not sensitive to the construction of the outcome measure, the study used a Rasch partial credit rating model to produce an alternative outcome measure and reestimate the model. (Appendixes D and E provide additional information on alternative specifications of the outcome measures.)

Alternative covariate specification

For the student impact analyses, the study considered different sets of individual- and school-level background variables to check whether the results were sensitive to the choice of covariates included in the estimation model. In particular, the benchmark model was compared with models estimated with a smaller number of covariates.

Results of alternative impact estimation for student outcomes

The alternatively scaled student outcome measure (z-scores), resulted in a larger effect size (0.379) than the benchmark model (table B.6). As with the benchmark model based on the linked scale scores, the estimated impact based on *z*-scores is statistically significant at the 5 percent level. The impact estimate based on a reduced set of covariates, excluding all individual-level variables and including only school-level baseline outcome measure and school size, is similar to that from the benchmark model (effect size of 0.204 versus 0.244) and statistically significant at the 5 percent level.

Alternative estimation methods yielded results comparable to the results of the benchmark model estimated by restricted maximum likelihood. The difference in the estimated average scale score between the treatment and control groups was 5.3–6.0 points across the alternative estimation methods. The average effect size was 0.208–0.234. (The overall effect size presented was computed as a weighted average of effect sizes of single-entity impacts, as opposed to the effect size of the weighted-average impact; the average effect sizes thus does not correspond proportionally to the average scale score difference.) Although there was some variation in the estimate, all alternative point estimates are well within one another's 95 percent confidence intervals, and all are statistically significant at the 1 percent level.

Table B.6 Sensitivity analyses for estimated impact of Pacific CHILD on student outcomes

Model	Weighted-average regression-adjusted means					Weighted average effect size
	Treatment schools	Control schools	Difference	Standard error	*p*-value	
Benchmark model						
Stanford 10 Achievement Test reading comprehension scale score, estimated by restricted maximum likelihood	634.3	629.0	5.3*	2.19	.017	0.244
Alternative scaling of outcome						
z-score	0.1	–0.1	0.1*	0.06	.023	0.379
Alternative estimation methods						
Maximum likelihood	634.1	628.1	6.0**	1.52	.000	0.234
Feasible generalized least squares with Swamy-Arora method	630.5	624.5	6.0**	1.92	.002	0.208
Generalized estimating equations with model-based standard error	636.5	630.4	6.0**	1.38	.000	0.216
Ordinary least squares, cluster-robust standard error	633.6	628.3	5.3**	1.12	.000	0.224
Alternative covariates						
Baseline score, school size, and blocks only	634.6	629.8	4.8*	2.18	.027	0.204

*Significant at the .05 level (two-tailed test), **significant at the .01 level (two-tailed test).

Note: The number of observations = 3 entities (3,052 students for the three entities combined). Scores are based on reading comprehension assessment data from the Stanford 10 Achievement Test (SAT 10) for American Samoa and the CNMI and the TerraNova for Hawai‘i. TerraNova scores were converted to SAT 10–equivalent scores using the methodology described in appendix E. For each entity, regression-adjusted means were computed at the means of the covariates; effect sizes were calculated by dividing the impact estimate by the standard deviation of the control group. Unless otherwise noted, each entity estimation included the following covariates: blocking variables, school-level baseline reading comprehension scale score, school size, student-to-teacher ratio, percentage of students eligible for free or reduced-price lunch, student gender, and student special education status. For generalized estimating equations, the student-to-teacher ratio and the percentage of students eligible for free or reduced-price lunch were excluded, because the model with the full set of covariates failed to converge. The overall impacts in scale score and effect size were computed as weighted means of the three single-entity impacts and the three corresponding effect sizes, with weights defined as the inverse of the variance of each scale score impact estimates.

Source: Authors’ analysis based on the student records collected for this study.

Results of alternative impact estimation for teacher outcomes

The estimation results for teacher knowledge and practice were also consistent across the estimation methods checked, as well as across alternative outcome measures and the alternative missing data adjustment method (tables B.7 and B.8). The alternative impact estimates were 0.34–0.39 for teacher knowledge and 0.63–0.64 for teacher practice. All alternative estimates are

statistically significant at the 5 percent level. These results are consistent with the findings based on the benchmark model reported in chapter 5 and are robust across methodological choices.

Table B.7 Sensitivity analyses of impact of Pacific CHILD on teacher knowledge assessment scores based on alternative estimation methods

Model	Weighted-average regression-adjusted means					Weighted average effect size	Number of observations (all entities total)
	Treatment schools	Control schools	Difference	Standard error	*p*-value		
Benchmark model (total knowledge assessment score estimated by restricted maximum likelihood)	27.00	25.04	1.96*	0.864	.023	0.35	190
Alternative scaling/specification of outcome							
Total score minus items with low item discrimination/item facility	26.76	24.86	1.89*	0.827	.022	0.34	190
Two-parameter item response theory–based measure	0.17	–0.17	0.34**	0.131	.009	0.39	190
Alternative missing data adjustment							
Dummy variable adjustment	26.97	25.00	1.97*	0.855	.021	0.35	197
Alternative estimation methods							
Feasible generalized least squares with Swamy-Arora method	27.02	25.02	2.00*	0.934	.032	0.36	190
Generalized estimating equations with model-based standard error	26.96	25.04	1.91**	0.655	.004	0.34	190
Ordinary least squares, cluster-robust standard error	26.96	25.05	1.91*	0.733	.012	0.34	190

*Significant at the .05 level (two-tailed test), **significant at the .01 level (two-tailed test).

Note: Regression-adjusted means were computed at the means of the covariates. Effect sizes were calculated by dividing the impact estimate by the standard deviation of the control group. For the teacher knowledge measure, the maximum likelihood estimation of the benchmark model failed to find a numerical solution.

Source: Authors' analysis based on student test records and teacher data collected for this study.

Table B.8 Sensitivity analyses of impact of Pacific CHILD on teacher practice observation scores based on alternative estimation methods

Model	Weighted-average regression-adjusted means					Weighted average effect size	Number of observations (all entities total)
	Treatment schools	Control schools	Difference	Standard error	*p*-value		
Benchmark model (total knowledge assessment score estimated by restricted maximum likelihood)	2.20	1.85	0.36**	0.121	.003	0.64	189
Alternative scaling of outcome							
Rasch-scaled measure	0.27	–0.14	0.40**	0.135	.003	0.64	189
Alternative missing data adjustment							
Dummy variable adjustment	2.20	1.84	0.36**	0.120	.003	0.64	198
Alternative estimation methods							
Maximum likelihood	2.20	1.85	0.36**	0.090	.000	0.63	189
Feasible generalized least squares with Swamy-Arora method	2.20	1.85	0.36*	0.146	.015	0.63	189
Generalized estimating equations with model-based standard error	2.20	1.85	0.36**	0.093	.000	0.64	189
Ordinary least squares, cluster-robust standard error	2.20	1.85	0.35**	0.092	.000	0.63	189

*Significant at the .05 level (two-tailed test), **significant at the .01 level (two-tailed test).

Note: Regression-adjusted means were computed at the means of the covariates. Effect sizes were calculated by dividing the impact estimate by the standard deviation of the control group. For the teacher knowledge measure, the maximum likelihood estimation of the benchmark model failed to find a numerical solution.

Source: Authors' analysis based on student test records and teacher data collected for this study.

Analyses of additional teacher outcome measures

The study team constructed the teacher outcome measures. (Chapter 3 and appendix D provide background information and detailed descriptions of these measures, including their limitations.) A concern regarding these study-developed measures is their content validity. This section presents the results of impact analyses using additional outcome measures based on the teacher data. The purpose of these analyses is to provide readers with additional information for evaluating concerns about content validity.

In the sensitivity analyses in the previous section, the alternative outcome measures examined were equivalent to the benchmark measure, except that they were specified or scaled differently based on the same sets of items (except for two items on the teacher knowledge assessment, which were excluded to improve reliability). This section examines additional outcome measures that were constructed using different sets of items, the content of which was not expected to be strictly equivalent to the benchmark measure. The analyses in this section thus do not provide a robustness test of the benchmark estimation results; they explore the validity of the instruments.

For the teacher knowledge assessment, concern was raised that the instrument may have been testing specific vocabulary rather than pedagogical knowledge.[107] To explore this concern, the study team constructed an outcome measure that excluded definitional questions and focused on application questions (table B.9).[108] As a reference, it also created another outcome measure based on definitional questions. The estimation results show a statically significant impact based on nondefinitional (application) items (effect size = 0.37, $p = 0.008$). The difference in impact based on definitional items is not statistically significant.

For the teacher practice, the study used average scores from a modified version of the Sheltered Instruction Observation Protocol (SIOP) instrument as the benchmark measure. The modified version used as the basis of the benchmark teacher practice measure included eight additional observation items covering areas targeted by the intervention. To explore whether the additional items may have biased the instrument to be overly aligned with the intervention, the study conducted impact estimates based only on the original SIOP (table B.9). The results show a statistically significant impact (effect size = 0.35, $p = .007$). As with additional knowledge measures, these additional impact estimation results are not intended to serve as a direct test of content validity; rather, they provide additional information that can be used to evaluate concerns about content validity.

---

[107] Another concern raised by external reviewers was that the instrument could be overtly aligned with the intervention. Appendix D describes how concerns about overalignment were addressed during instrument development. It also discusses the results of item analyses conducted to examine differential response patterns between treatment and control groups (differential item functioning) to explore whether items systematically favored treatment group teachers over control group teachers.

[108] Definitional questions are questions that test knowledge of what a particular term means ("What does __ mean?"). Fifteen questions (questions 3, 8, 10, 12, 13, 16, 18, 19, 21, 22, 24, 26, 32, 36, and 37) were identified as nondefinitional (application based) ("What are examples of ___?").

Table B.9 Estimated impact of Pacific CHILD on teacher knowledge and practice based on alternative outcome measures

Outcome measure	Weighted-average regression-adjusted means					Weighted average effect size	Number of observations (all entities total)
	Treatment schools	Control schools	Difference	Standard error	*p*-value		
Teacher knowledge							
Benchmark outcome measure (total knowledge assessment score)	27.00	25.04	1.96*	0.864	.023	0.35	190
Other outcome measures							
Application (non-definitional) item total score	11.72	10.73	0.99**	0.375	.008	0.37	190
Definitional item total score	15.27	14.32	0.95	0.585	.104	0.27	190
Teacher practice							
Benchmark outcome measure (average practice observation score)	2.20	1.85	0.36**	0.121	.003	0.64	189
Other outcome measures							
Average practice observation score, based on original 30 Sheltered Instruction Observation Protocol (SIOP) items	2.41	2.06	0.35**	0.128	.007	0.57	189

*Significant at the .05 level (two-tailed test), **significant at the .01 level (two-tailed test).

Note: Regression-adjusted means were computed at the means of the covariates. Effect sizes were calculated by dividing the impact estimate by the standard deviation of the control group. Impact estimation for all but one outcome measure was based on the same covariates as the benchmark model and on restricted maximum likelihood (for the nondefinitional item measure of teacher knowledge, the feasible generalized least squares estimate based on the full set of covariates is reported). The numerical solution to the restricted maximum likelihood estimation was found when excluding three covariates (years at current school, percent of students eligible for free or reduced-price lunch, and student-to-teacher ratio) (impact estimate = 0.92, standard error = 0.368, p = .01). The benchmark average practice score was computed based on 37 items, excluding one item that was not rated consistently. The average practice observation score based on original SIOP items excluded the same item and was calculated based on 29 items.

Source: Authors' analysis based on student test records and teacher data collected for this study.

Appendix C: Technical notes on exploratory analyses and results of analysis of impact on teachers in non-Hawai‘i subsamples

This appendix provides technical details on the exploratory analyses reported in chapter 6. It also presents the analyses of impacts on teachers in the non-Hawai‘i subsamples.

Analytic approach

In contrast to the confirmatory analyses presented in chapter 5, which tested specific hypotheses derived from the underlying theoretical model of the intervention, the exploratory analyses were based on research questions that aimed to uncover patterns of impacts and develop potential hypotheses for further investigation.[109] These questions covered four areas of inquiry: patterns of impacts in the Hawai‘i subsample, patterns of impacts for subscales of the teacher practice measure, patterns of impacts moderated by teacher characteristics, and patterns of impacts in the non-Hawai‘i subsamples.[110] (Chapter 6 presents the findings on the first three areas of inquiries and on impact estimates for students in non-Hawai‘i entities; this appendix presents the findings on impact estimates for teachers in the non-Hawai‘i entities.)

Analytic methods used in the exploratory analyses were parallel to those used in the confirmatory impact analyses. In the exploratory analyses, however, statistical testing was used to help identify patterns rather than test explicit hypotheses based on a theoretical model.

Like the confirmatory impact analyses, the exploratory analyses used the hierarchical linear modeling presented in appendix B. For analysis of impacts for entity subsamples, the estimation model specified was the same as the confirmatory model, except for entity-specific fixed effects. Using the same notation used in appendix B, this model can be expressed in reduced form as:

$$Y_{ij}^{k} = \gamma_0 + \gamma_1 (STATUS)_j + \sum_{q=1}^{Q} \beta_q X_{qij}^{k} + \sum_{s=2}^{S} \gamma_s W_{sj}^{k} + \varepsilon_{ij} + u_j \qquad \text{(Equation C.1)}$$

where Y^k_{ij} is the observed outcome for student or teacher i in school j in entity k; X^k_{qij} is the qth individual-level covariate for student or teacher i in school j in entity k; W^k_{sj} is the sth school-level covariate for school j in entity k; ε_{ij} is the residual term specific to teacher or student i in school j; and u_j is the residual specific to the jth school, representing the unobserved random school effect. *STATUS* is a variable indicating whether school j was randomly assigned to receive Pacific CHILD (*STATUS* = 1) or not (*STATUS* = 0). As in the confirmatory analysis, the parameter of interest was γ_1, which captures the effect of a randomized offer of Pacific CHILD

[109] All exploratory analyses presented were conducted in an experimental framework. Nonexperimental analyses are not discussed in this report.

[110] This study was designed at the outset with a sufficiently large Hawai‘i subsample, with the view to testing the hypothesis that Pacific CHILD had impacts on students and teacher in Hawai‘i. The Hawai‘i subgroup analysis is presented as an exploratory inquiry because it was not designed to address the confirmatory research questions regarding the effectiveness of Pacific CHILD across the entities. The primary purpose of this inquiry was to provide supplementary findings to the core investigation based on the pooled sample.

on student and teacher outcomes in a given entity k. Exploratory testing of γ_1 was analogous to that in the confirmatory analysis: an the estimate of γ_1 that was statistically differed from zero at the 5 percent significance level with a two-tailed test was taken as an indication that the intervention had impact. This analysis did not test whether a pre-established null hypotheses about γ_1 could be refuted; rather, it examined the potential for establishing such a confirmatory hypothesis for γ_1.

For exploratory analyses of impacts on the subscales of the teacher practice measure, each subscale was examined independently, applying the same hierarchical linear modeling model used in the confirmatory analysis but with entity-specific fixed effects. For analyses of the moderating effects of teachers' characteristics on impacts, the reduced-form hierarchical linear modeling model was expanded for a given entity k (Hawai'i in this case):

$$\begin{aligned} Y_{ij}^{k} = \gamma_0 + \gamma_1 (STATUS)_j + \sum_{q=1}^{Q} \beta_q X_{qij}^{k} + \sum_{s=2}^{S} \gamma_s W_{sj}^{k} \\ + \lambda_{EXP} (EXP)_i (STATUS)_j + \lambda_{MA} (MA)_i (STATUS)_j \\ + \beta_{EXP} (EXP)_i + \beta_{MA} (MA)_i + \varepsilon_{ij} + u_j \end{aligned} \qquad \text{(Equation C.2)}$$

where EXP is an indicator for years of education; MA is a binary indicator for completing an advanced degree; and λ_{EXP} and λ_{MA} are the parameters measuring the fixed experience- and education level–specific effects that moderate the impacts on teacher outcomes. The exploratory analyses investigated whether each of these coefficients on interaction terms (λ_{EXP} and λ_{MA}) was statistically significant. To explore the differential impacts of teachers' experience and their level of education based on the pooled data, the study expanded the model with entity-specific fixed effects (see equation B.4 in appendix B).

As in the confirmatory analysis, the restricted maximum likelihood estimator was used as the primary estimation method. Missing outcome data were addressed by listwise deletion, and missing covariates were addressed by applying the dummy variable adjustment method. The minimum detectable effect size estimated a priori for each question is summarized in table C.1.

The following sections present details on the investigation of each exploratory question, including descriptions of the benchmark model specifications reported in chapter 6 and alternative models estimated.

Table C.1 Minimum detectable effect size for analysis of exploratory questions on impact of Pacific CHILD on students and teachers

Exploratory research question number	Outcome	Minimum detectable effect size[a]
1	Student outcome in Hawai‘i (cross-sectional subsample)	.22
2	Student outcome in Hawai‘i (longitudinal subsample)	.24
3–4	Teacher outcomes in Hawai‘i	.61
5	Student outcome in American Samoa	.61
6	Student outcome in the Commonwealth of the Northern Mariana Islands	.41
7–10	Teacher practice observation subscales	.46
11–14	Moderating effects on teacher outcomes (pooled data)	.46
	Moderating effects on teacher outcomes (Hawai‘i subsample)	.61
15–16	Teacher outcomes in non-Hawai‘i subsample	.71

a. Calculated at power =.8 and significance level =.05, assuming that intraclass correlation coefficient =.05, R^2 = .5 for student outcomes, and R^2 = 0.0 for teacher outcomes.

Note: The subsample analyses of student outcomes for Hawai‘i were conducted using two alternative definitions of the cohorts, “cross-sectional” and “longitudinal” subsamples. The cross-sectional sample included students who happened to be at study schools at the end of the two-year intervention. The longitudinal sample included students who were at the study schools from the start through the end of the intervention. The longitudinal subsample consisted of students who could have been exposed to the two full years of the intervention, from the start to the end; in contrast with the cross-sectional subsample of students who were at schools that could have been exposed to the two years of the intervention.

Source: Authors’ calculations using Optimal Design software.

Analysis of Hawai‘i and non-Hawai‘i subsamples

This section provides additional information on the analyses by entity.

Subsamples

The subsamples used for the analyses of teachers in Hawai‘i and students and teachers in the other entities were subsets of the corresponding confirmatory impact analysis samples. As such, they represented individuals at the study schools near the end of two years of implementation of Pacific CHILD.

The subsample analyses of student outcomes for Hawai‘i were conducted using two alternative definitions of the cohorts, “cross-sectional” and “longitudinal” subsamples (table C.2). The “cross-sectional” sample included students who happened to be at study schools at the end of the two-year intervention“. The longitudinal” sample included students who were at the study schools from the start through the end of the intervention.

The "cross-sectional" subsample was conceptually equivalent to the impact analysis sample, defined as grade 5 students who were at study schools at the time of data collection near the end of the two-year implementation. The "longitudinal" subsample was a subset of the cross-sectional sample, defined as grade 5 students who were present at the school as grade 3 students at the time of random assignment and as grade 5 students at the time of data collection two years later. This longitudinal subsample consisted of students who could have been exposed to the two full years of the intervention, from the start to the end, in contrast with the cross-sectional subsample of students who were at schools that could have been exposed to the two years of the intervention. The cross-sectional subsample thus included students who transferred to the study schools after random assignment and could not have been exposed to the full two-year intervention. Because of the limitations of the data, it was not possible to construct equivalent longitudinal cohorts for teachers or for students in other entities. (For the characteristics of teachers by entity, see table 2.5 in chapter 2.)

Table C.2 Profiles of student samples used to assess impact of Pacific CHILD

(percent, except where otherwise indicated)

	Hawai'i			
Student characteristic	**Cross-sectional subsample**	**Longitudinal subsample**	**American Samoa**	**Commonwealth of the Northern Mariana Islands**
Female	48.1	46.9	42.7	47.8
Race/ethnicity other than White	83.5	85.4	100.0	99.7
Special education student	11.1	11.1	13.5	8.4
English language learner student	13.8	12.2	—	—
Eligible for free or reduced-price lunch	56.7	55.8	—	—
Mean SAT 10–equivalent score at baseline	606.9	606.9	—	—
Number of students	2,175	1,741	185	692

— is not available. SAT 10 is the Stanford 10 Achievement Test.

Note: The cross-sectional subsample for Hawai'i and the subsamples for American Samoa and the Commonwealth of the Northern Mariana Islands consisted of grade 5 students who were at the study schools at the time of outcome data collection. The longitudinal subsample for Hawai'i consisted of grade 5 students who were at the study schools at baseline and at the time of outcome data collection. For the cross-sectional subsample of Hawai'i, the mean baseline SAT 10–equivalent scores were calculated based on 1,741 students for whom base year test records were available (see appendix E for the conversion methodology used).

Source: Authors' analysis based on student records provided by the American Samoa Department of Education, the Commonwealth of the Northern Mariana Islands Public School System, and the Hawai'i Department of Education.

Data and treatment of missing data

All exploratory analyses used the same student outcome measure used in the confirmatory analyses, namely, the Stanford 10 Achievement Test (SAT 10) (and SAT 10–equivalent) reading comprehension subscores for grade 5 students. For the exploratory teacher analyses, outcome measures based on the teacher knowledge assessments and teacher practice observations were used, including subscales not investigated in the confirmatory analyses (table C.3).

Table C.3 Covariates included in exploratory analyses of impact of Pacific CHILD in Hawai‘i and non-Hawai‘i student and teachers subsamples

	Hawai‘i		Non-Hawai‘i	
Level/variable	**Student sample**	**Teacher sample**	**Student sample**	**Teacher sample**
School-level				
Assignment condition: treatment group indicator (binary indicator)	✓	✓	✓	✓
Assignment block and entity indicators (binary indicator)[a]	✓	✓	✓	✓
Baseline average reading comprehension subtest score[b]		✓	✓	✓
School size (total number of students in school) at baseline	✓	✓	✓	✓
Student-to-teacher ratio at baseline	✓	na	✓	na
Percent of students eligible for free or reduced-price lunch at baseline	✓	na	✓	na
Percent of English language learner students at baseline	✓	na		na
Teacher-level				
Grade taught (binary indicator)		✓		✓
Gender (binary indicator)		✓		✓
Primary language not English (binary indicator)		✓		✓
Total years teaching		✓		✓
Total years teaching at current school		✓		✓
Completed less than a bachelor’s degree (binary indicator)				✓
Completed more than a bachelor’s degree (binary indicator)		✓		✓
Certification status (binary indicator)				✓
Race/ethnicity (binary indicators for White, Asian, and Pacific Islander)		✓		✓
Student-level				
Gender (binary indicator)	✓		✓	
Race/ethnicity (binary indicator for non-Asian/non–Pacific Islander)			✓	
Race/ethnicity (binary indicators for Filipino, Hawai‘i an, Samoan, East Asian, other Asian, and non-Asian/non-White)[c]	✓			
Special education status (binary indicator)	✓		✓	
English language learner (binary indicator)	✓			
Eligible for free and reduced-price lunch (binary indicator)	✓			
Baseline reading comprehension subtest score (grade 3 scores)	✓			

a. Because schools were blocked within entities, a noncollinear subset of these indicators was included in the estimation.

b. School-level baseline scores for the target grades were computed based on individual-level Stanford 10 Achievement Test (SAT 10) data. Grade 5 scores were not available for American Samoa; grade 4 scores were not available for the Commonwealth of the Northern Mariana Islands (CNMI) at baseline. For student outcome analyses, the baseline year school average was computed based on grade 5 scores for schools in Hawai‘i and the CNMI and on grade 4 scores for schools in American Samoa. For Hawai‘i, TerraNova scores were converted to equivalent SAT 10 scores, using the methodology described in appendix E. For teacher outcome analyses, the baseline year school average for Hawai‘i was computed across the two grades. For the Hawai‘i subsample, maximization routines for some restricted maximum likelihood models (for exploratory questions 3, 12, and 15) failed to attain a local maximum when the school-level baseline test score was included. In these cases, the results estimated without school-level baseline reading test scores are reported.

c. Included only in the estimation for the Commonwealth of the Northern Mariana Islands sample.

Source: Authors' analysis of student records provided by the American Samoa Department of Education, the Commonwealth of the Northern Mariana Islands Public School System, and the Hawai‘i Department of Education.

For analyses of the Hawai‘i student subsample, additional covariates were added to improve the precision of the impact estimates. These covariates included English language learner status, eligibility for free or reduced-price lunch, additional race/ethnicity categories, and individual-level (rather than school-level) baseline reading comprehension scores.[111] None of the records with outcome data was missing student characteristics data, except baseline test data for the cross-sectional sample. The dummy variable adjustment method was applied to address the missing baseline individual-level test scores for the cross-sectional subsample of students. Missing records were replaced with the sample mean and included in estimation along with the dummy variable indicating records with imputed scores. The "longitudinal" subsample of Hawai‘i students was equivalent to the "cross-sectional" subsample of students (after the listwise deletion of records with missing baseline data).[112]

For the Hawai‘i subsample of teachers, missing covariates were minimal (one record for the teacher knowledge estimation and two records for the teacher practice estimation were missing). Listwise deletion was used to address these missing covariates.[113]

For the non-Hawai‘i student subsamples, listwise deletion was used to address missing outcome data. The covariates used in the non-Hawai‘i subsamples were the same as those used in the confirmatory analyses. No covariates were missing for students with outcome data.

For the non-Hawai‘i subsample analysis of teacher outcomes, the estimation models included the same set of individual-level covariates, except for those excluded because of perfect collinearity. The number of school-level covariates was minimized to reduce the loss of degrees of freedom, which was a concern because of the relatively small size of the teacher subsamples ($N \leq 117$). The school-level covariates were pared down, based on stepwise likelihood ratio tests of the contribution of each covariate to the model.[114] The dummy variable adjustment method was used to address missing covariates in the teacher subsample.

---

[111] As a sensitivity analysis, the study also estimated the model using exactly the same set of covariates used in the benchmark model for the confirmatory analyses in chapter 5.

[112] This equivalence exists because the baseline test data were used to identify the students enrolled in the study schools at baseline.

[113] The estimate using the dummy variable adjustment produced the same results as the estimate without the adjustment.

[114] Likelihood ratio tests were based on models without other covariates, except for assignment condition and block indicators. For the teacher subsamples, school-level covariates were pared down to two (school size and baseline test scores), based on a pairwise process. However, for some fully specified restricted maximum likelihood estimation models, maximization routines failed to attain local maximums. In these cases, models excluding one additional school-level variable (baseline test score) were estimated, and alternative estimation methods with closed-form solutions with a full set for covariates were estimated to check that the impact estimates were consistent across estimation methods and covariate specifications. For teacher-level outcomes, results based on restricted maximum likelihood are reported, unless otherwise noted.

Additional considerations for model specifications and estimation methods for analyses of student outcome data in the Hawai‘i subsample

The analyses of the Hawai‘i student subsamples included the individual-level baseline test score as a covariate. Both estimated SAT 10–equivalent and TerraNova scale scores were used as the baseline reading achievement covariate. The results based on both measures yielded very similar estimates (table C.4). The results reported in chapter 6 were based on the SAT 10–equivalent scores as the baseline test measure. For the Hawai‘i student subsample analyses, an alternative model that included dummy variables for schools with a high rate of missing baseline test scores (more than 40 percent) was also explored.[115] The rate of missing baseline test data in these schools was likely high because of structural reasons independent of the intervention (for example, the location of a school in a highly mobile community). The indicators for these schools were included to control for any systematic difference in the mean scores. The results showed that inclusion of these school-level low response indicators increased the impact estimates for the cross-sectional subsample (effect size = 0.05).

For the cross-sectional Hawai‘i subsample (N = 1,741), the fully specified models were not successfully estimated with the restricted maximum likelihood method.[116] The fully specified model, with additional covariates for Hawai‘i, was therefore estimated using the alternative estimation methods discussed in appendix B. All alternative estimation methods yielded consistent estimation results. The results reported in chapter 6 were based on the fully specified model estimated using the Swamy-Arora feasible generalized least squares estimator.

The results from the by-entity subsample analyses point to several directions for future investigation, including analysis of differences in implementation of the intervention across entities and studies of differential impacts across entities.

[115] Baseline test scores were missing for more than 40 percent of the students at fewer than four schools in Hawai‘i.

[116] The attempts to maximize numerically the likelihood functions for the restricted maximum likelihood (and maximum likelihood) estimators for the fully specified models failed under alternative maximization algorithms.

Table C.4 Estimated impact of Pacific CHILD on reading comprehension of grade 5 students in Hawai‘i , alternative covariates

	Regression-adjusted means						
Covariates included/estimation methods	**Treatment schools**	**Control schools**	**Difference**	**Standard error**	***p*-value**	**Effect size**	**Number of observations**
Cross-sectional subsample: Students at study schools in second year							
Benchmark model without Hawai‘i -specific covariates (same as chapter 5)	640.0	636.0	4.02	2.50	0.108	0.10	2,175
Hawai‘i -specific covariates with SAT 10–equivalent baseline score	640.0	636.0	4.04*	1.94	0.037	0.10	2,175
Hawai‘i -specific covariates with TerraNova baseline score	640.0	636.0	4.05*	1.96	0.039	0.11	2,175
Hawai‘i -specific covariates with SAT 10–equivalent baseline score and low-response school indicators	640.9	635.1	5.81*	2.67	0.029	0.15	2,175
Longitudinal subsample: Students at study schools at baseline and in second year							
Benchmark model without Hawai‘i -specific covariates (same as chapter 5), restricted maximum likelihood	641.6	635.2	6.35*	2.73	0.020	0.17	1,741
Hawai‘i -specific covariates with SAT 10–equivalent baseline score, Swamy-Arora feasible generalized least squares	640.4	636.4	4.03**	1.51	0.008	0.11	1,741
Hawai‘i -specific covariates with SAT 10–equivalent baseline score, generalized estimating equations	640.7	636.1	4.57**	0.81	0.000	0.12	1,741
Hawai‘i -specific covariates with SAT 10–equivalent baseline score, ordinary least squares with cluster robust errors	640.4	636.4	4.03**	0.98	0.000	0.11	1,741

SAT 10 is the Stanford 10 Achievement Test.

*Significant at the .05 level (two-tailed test). **Significant at the .01 level (two-tailed test).

Note: Regression-adjusted means were computed at the means of the covariates. The estimates based on the restricted maximum likelihood method are reported in the table, unless otherwise specified. Effect sizes were calculated by dividing the impact estimate by the standard deviation of the control group. TerraNova scores were converted to SAT 10–equivalent scores using the methodology described in appendix E.

Source: Authors’ calculations based on data collected for this study.

Analyses of subscales of teacher practice measure

For analyses of subscales of the teacher practice measure, the study estimated the model based on the pooled data and using alternative outcome measures to represent different subsets of the classroom practice measure. Summary statistics for subscales used as alternative outcome measures are provided in table C.5. As in the confirmatory analyses, the impacts on the teacher practice subscales were estimated based on a hierarchical linear modeling approach to account for the nested nature of the data. The same set of covariates used in the confirmatory analyses was included in the estimation. Missing outcome data were addressed by listwise deletion; missing covariates were addressed by applying the dummy variable adjustment method. The results of the analyses for subscales with a Cronbach's alpha of .70 or higher are presented in chapter 6 (table 6.5). The results of the analyses for subscales with a Cronbach's 'alpha of less than .70 are provided in table C.6

Table C.5 Summary of subscale measures of teacher practice measure

Dimension/subarea	Number of items	Cronbach's alpha	Mean score	Standard deviation	Minimum score	Maximum score
Preparation dimension	6	.72	2.08	.69	0.0	3.6
Instruction dimension[a]	19	.87	2.22	.63	0.9	3.4
Building background (subarea)	3	.33	1.70	.93	0.0	4.0
Comprehensible input (subarea)	3	.57	2.56	.63	1.0	3.7
Strategies (subarea)	3	.76	2.15	.93	0.0	4.0
Interaction[a] (subarea)	3	.55	2.00	.82	0.0	4.0
Practice and application (subarea)	3	.66	2.20	.85	0.0	4.0
Lesson delivery (subarea)	4	.70	2.60	.75	1.0	4.0
Review and evaluation dimension	4	.63	1.56	.80	0.0	4.0
Additional items: strategies emphasized by the intervention	8	.59	1.15	.59	0.1	2.9

a. One item that was inconsistently rated was excluded in computing the subscale score.

Source: Authors' calculations based on data collected for this study.

Table C.6 Estimated impact of Pacific CHILD on teacher practice subscales with Cronbach's alpha less than .70

Subscale component[a] (corresponding exploratory question number)	Cronbach's alpha	Regression-adjusted mean of average rating of teacher practice observation subscale					Effect size
		Treatment schools	Control schools	Difference	Standard error	*p*-value	
Subareas of instruction dimension (exploratory question 8)							
Building background	.33	1.75	1.71	.04	.180	.846	.04
Comprehensible input	.57	2.85	2.53	.32*	.136	.017	.52
Interaction	.55	2.41	1.95	.45**	.130	.000	.55
Practice and application	.66	2.58	2.14	.44*	.181	.014	.52
Review and evaluation dimension (exploratory question 9)	.63	1.97	1.52	.46**	.156	.004	.57
Additional section on reading and instructional strategies (exploratory question 10)	.59	1.49	1.10	.40**	.134	.003	.67

*Significant at the .05 level (two-tailed test), **significant at the .01 level (two-tailed test).

Note: Regression-adjusted means were computed at the means of the covariates. Effect sizes were calculated by dividing the impact estimate by the standard deviation of the control group. The hierarchical linear modeling models corresponding to exploratory questions 9–11 were estimated using the restricted maximum likelihood method. Number of observations = 189. For the results of the analyses for subscales with a Cronbach's alpha of .70 or higher, see table 6.5 in chapter 6.

Source: Authors' calculations based on data collected for this study.

Analyses of moderating effects of teacher characteristics

The study explored the moderating effects of teacher experience and education by adding interaction terms to the model. The treatment indicator was multiplied by years of teaching and highest education status to create the interaction terms. No predictions were made for the direction of moderating effects before the estimation. Although it might be expected that teachers with less education or experience would gain more from the intervention (because the intervention could help them acquire new skills and knowledge), it is also plausible that teachers with more education or more experience would gain more from the intervention (because such teachers are better equipped to understand and adopt the contents of the intervention). The purpose of this exploratory analysis was to examine the potential presence and direction of moderating effects.

The study estimated the coefficients of interaction terms based on the pooled data and the Hawai‘i teacher subsample. Hawai‘i data were used to provide within-entity estimates of

moderating effects. (Teacher samples from American Samoa and the CNMI were not large enough to conduct reliable analysis by entity.) The Hawai‘i teacher subsample analysis also addressed the concern that interaction terms estimated with pooled data might not accurately reflect the effects of teacher characteristics because of potential correlations between teacher background characteristics and entity.[117] Analysis based on the pooled data was conducted as the preferred approach, despite the concern noted above, because the pooled data afforded more power and because preliminary specification tests rejected the assumption of entity-specific slopes on the treatment indicator (see appendix B), reducing concerns about correlations between teacher background characteristics and entities in estimating moderating effects.

These differential impacts on teacher outcomes by teacher characteristics were estimated based on a hierarchical linear modeling model using the same outcome measures used in earlier analyses. Listwise deletion was applied to handle missing data. Total years of teaching at baseline and an indicator for an advanced degree were examined as measures of moderating effects. The estimation results based on the pooled data are provided in chapter 6, and the results based on the Hawai‘i subsample are provided in table C.7.

Table C.7 Effect of additional year of experience or advanced degree on estimated impact of Pacific CHILD on Hawai‘i teacher outcomes

Variable	Marginal effects on impact	Standard error	*p*-value	Number of observations
Knowledge: Total score on knowledge assessment				
Additional year of experience	.32**	.123	.009	116
Advanced degree (versus bachelor's)	–.13	1.752	.942	116
Teacher practice: Average score on classroom observation				
Additional year of experience	.00	.014	.986	117
Advanced degree (versus bachelor's)	.05	.182	.793	117

*Significant at the .05 level (two-tailed test), **significant at the .01 level (two-tailed test).

Note: Regression-adjusted means were computed at the means of all covariates except education, which was set at the bachelor's level. The parameters of the two-level hierarchical linear modeling model were estimated using the restricted maximum likelihood method. Effect sizes were calculated by dividing the impact estimate by the standard deviation of the control group. For the teacher knowledge estimation, the school-level baseline reading comprehension score was excluded as a covariate. When the variable was included, maximization routines for restricted maximum likelihood failed to attain a local maximum. Alternative estimators yielded estimates comparable in size and significance to those reported in the table. For each outcome measure, the moderating effects of years of experience and of education were estimated jointly in a single estimation model (see equation B.6).

Source: Authors' calculations based on data collected for this study.

The results were consistent regardless of whether the pooled sample or the Hawai‘i subsample was used. The impacts on teacher outcome measures did not vary with teachers' experience or education, except for the moderating effect of years of teaching on teacher knowledge. Based on

[117] For example, this concern applied in the case of teachers whose highest education level was less than a bachelor's degree. Teachers who had less than a bachelor's degree were concentrated in one entity.

the initial findings, additional specifications allowing discontinuous slopes on the interaction terms were examined to further explore the relationship between teaching experience and the impact of Pacific CHILD on teacher knowledge. The estimation from the piecewise interaction terms indicates that the positive moderating effects are statistically significant for teachers with 10 or more years of experience (table C.8). Future studies are needed to further investigate the potential moderating effect of teacher experience on the impacts of the intervention, including the nonlinearity of such effects.

Table C.8 Estimates of moderating effects of additional year of teaching experience on impact on total score of teacher knowledge assessment

Variable	Marginal effects on impact	Standard error	*p*-value	Number of observations
Pooled data				
0–4 additional years	.45	.523	.391	190
5–10 additional years	.30	.189	.113	190
11 or more additional years	.23*	.099	.022	190
Hawai'i subsample				
0–4 additional years	1.14	.706	.107	116
5–10 additional years	.47	.257	.065	116
11 or more additional years	.39**	.134	.004	116

*Significant at the .05 level (two-tailed test), **significant at the .01 percent level (two-tailed test).

Note: Regression-adjusted means were computed at the means of the covariates except for education, which was set at the bachelor's degree level. The parameters of the two-level hierarchical linear modeling model were estimated using the restricted maximum likelihood method. For the estimation conducted with the Hawai'i subsample, the school-level baseline reading test score was excluded from covariates, because maximization routines failed to attain a local maximum when the school-level baseline test score was included.

Source: Authors' calculations based on data collected for this study.

Analyses of impact on teachers in American Samoa and the Commonwealth of the Northern Mariana Islands

This section presents analyses on teachers conducted for the non-Hawai'i entities. The study team determined that the teacher impacts could not be reliably estimated separately by entity because of the small subsample size (see table C.1).[118] Only the results based on the two non-Hawai'i entities combined are therefore presented below. The guiding questions explored were as follows:

- Exploratory question 15: Did grade 4 and grade 5 teachers at schools in American Samoa and the CNMI that were offered Pacific CHILD for two years perform differently on assessments of their knowledge of theories and strategies related to effective reading instruction,

[118] The teacher subsamples consisted of 18 teachers in American Samoa (effect size = 1.58) and 62 in the CNMI (effect size = 0.92). The effect size for the combined teacher sample was 0.71.

including English language learner students-focused theories and strategies, than grade 4 and grade 5 teachers at schools in American Samoa and the CNMI that were not offered Pacific CHILD?

- Exploratory question 16: Did grade 4 and grade 5 teachers at schools in American Samoa and the CNMI that were offered Pacific CHILD for two years perform differently on assessments of their instructional practice for enhancing student reading comprehension, including English language learner-focused practice, than grade 4 and grade 5 teachers at schools in American Samoa and the CNMI that were not offered Pacific CHILD?

Each of these questions was investigated independently, based on the same hierarchical linear model specified for the confirmatory analysis. The same outcome measures—total scores from the teacher knowledge assessment and average scores from the teacher observation—were used.

The result based on the sample combining American Samoa and the CNMI are provided in table C.9. The estimated difference in the total teacher knowledge assessment scores between teachers at treatment and control group schools is 3.84 points and is statistically significant (effect size = 0.69, p = .009). The estimated impact on the average classroom practice observation scores is not statistically significant.

Table C.9 Estimated impact of Pacific CHILD on knowledge and practice of grade 4 and grade 5 teachers in American Samoa and the Commonwealth of the Northern Mariana Islands (combined)

Outcome measure (corresponding exploratory question number)	Regression-adjusted means					Effect size	Number of observations
	Treatment schools	Control schools	Difference	Standard error	*p*-value		
Total score on knowledge assessment (exploratory question 15)	24.7	2.9	3.84**	1.462	.009	.69	80
Average score on classroom observation (exploratory question 16)	1.98	1.68	.31	.185	.096	.54	79

**Significant at the .01 level (two-tailed test).

Note: Regression-adjusted means were computed at the means of the covariates. Effect sizes were calculated by dividing the impact estimate by the standard deviation of the control group. The hierarchical linear modeling models were estimated using the restricted maximum likelihood method.

Source: Authors' calculations based on data collected for this study.

Appendix D: Teacher outcome measures

This appendix provides additional information on the selection and properties of the instruments used for measuring teacher outcomes.

Measuring teacher knowledge: The teacher knowledge assessment

The teacher knowledge assessment developed for this study was used to measure teacher knowledge of pedagogical theories and classroom practices, including theories and practices that address English language learner students in the classroom.

Format and development of teacher knowledge assessment

Toward the end of the first and second years of the intervention, teachers completed the teacher knowledge assessment, a 40-item, 5-option multiple-choice assessment created for this study. The assessment measures general knowledge of pedagogical theories and practices for classroom instruction in the following areas: English reading instruction, second language acquisition theory, instructional methodologies for teaching English as a second language, theories of cognition, principles of lesson planning, assessment and grouping strategies, and scaffolding techniques. It includes both concepts that were expected to be emphasized in Pacific CHILD (26 items) and concepts that may not be emphasized in Pacific CHILD (14 items). The items that reflect concepts from Pacific CHILD were developed from the Pacific CHILD manuals and the general pedagogical literature that Pacific CHILD is based on, including research on differentiated instruction, interactive tasks, word parts, and vocabulary.[119] It was thus expected that well-trained teachers would have been familiar with these strategies and able to score well on concepts emphasized in Pacific CHILD, whether they were in treatment or control group schools. It was not expected that all teachers participating in Pacific CHILD would necessarily score well on these items—that is, the items that reflect Pacific CHILD content were designed so that they would not simply discriminate between Pacific CHILD participants and nonparticipants. The items that reflect concepts not specifically emphasized in Pacific CHILD were developed from theories and practices covered in undergraduate and graduate coursework, including research on scaffolding, additive bilingualism, and comprehensible input. Well-trained teachers in both groups were expected to score higher on these items than teachers who were not well trained.

The assessment was piloted with a small sample of teachers in the Pacific region. Based on analyses of data collected toward the end of the first year of the intervention, the study team replaced five items and revised another five. The revised version of the assessment used in the second year retained the same number of items (40) and the same distribution of general knowledge items and items relevant to Pacific CHILD. The reliability of the assessment rose from .76 at the end of the first year to .78 at the end of the second year, as measured by the

[119] As described in chapter 1, the content of Pacific CHILD is not unique to Pacific CHILD but is based on accepted, research-based reading comprehension and instructional strategies covered in undergraduate and graduate coursework and teacher professional development activities.

Kuder-Richardson formula 20 (KR-20).[120] The data collected toward the end of the second year of the intervention were used to measure the impact of Pacific CHILD on teacher knowledge.

To explore the possibility that the assessment systematically favored the treatment group, item-level statistics were compared for teachers at treatment and control group schools who completed the teacher knowledge assessment toward the end of the second year. Item facility (the proportion of teachers who answer an item correctly) was compared across the conditions to check whether any of the items were answered correctly only or mostly by teachers at treatment group schools.[121] The difference in item facility ranged from –.11 (.36 for the treatment group and .47 for the control group) to .26 (.87 for the treatment group and .61 for the control group). No items were answered correctly only or mostly by teachers at treatment group schools. The average difference in item facility was .03.

The study also conducted differential item functioning (DIF) analyses, to examine whether items were fair for teachers at treatment and control group schools with the same estimated ability level. Mantel–Haenszel statistics were calculated using Winsteps and interpreted following the guidelines of the Educational Testing Services for DIF categories (A = negligible (DIF < 1.0); B = slight to moderate (1.0 < DIF < 1.5); C = moderate to large (DIF > 1.5) (see Zieky 2003). Five items exhibited category B or C levels of differential item functioning ($p < .05$), with three items (questions 8, 19, and 23) favoring teachers at treatment group schools and two items (questions7 and 9) favoring teachers at control group schools. Given these results, it was assumed that the knowledge assessment was not an unfair test that overly favored teachers at treatment group schools.

Constructing the outcome measure for teacher knowledge

The benchmark outcome measure for teacher knowledge was calculated as a total score, the number of correct responses on the teacher knowledge assessment. Two alternative outcome measures were constructed for sensitivity analyses. The first was an alternative total score excluding poorly performing items and ability measures using an item response theory model. The alternative total score was computed as the total number of correct responses, excluding two items with low item facility and low item discrimination. (Item facility refers to the overall difficulty of an item; item discrimination refers to the extent to which an item distinguishes between high- and low-performing examinees.) The criteria for excluding the two items were item facility below .20 (that is, 20 percent or less of examinees answered the item correctly) and item discrimination below .20 (considered a cut-off for marginal items, where item discriminations of .19 and below are considered weak, following Ebel and Frisbie 1991). The resulting alternative measure had a reliability of .79, as measured by the KR-20 coefficient.

The other alternative outcome measure was constructed using a two-parameter item response theory model with the software program BILOG-MG.[122] Item response theory models produce ability measures (referred to as *person locations* or *thetas*) along a common scale by modeling

---

[120] The Kuder-Richardson formula 20 (KR-20) is a measure of internal consistency reliability for assessments with dichotomous scoring (for example, correct–incorrect). The KR-20 reliability estimate depends on the range and difficulty of items, the spread in test-taker scores, and the length of the assessment.

[121] Some test developers refer to item facility as the *p-value* (percent correct) for an item.

[122] A two-parameter model was used because the three-parameter item-response theory model did not converge for the sample of teachers ($N = 197$).

the probability that an examinee will correctly answer a specific item given his or her overall ability and the characteristics of the item. A two-parameter model takes into account both the item difficulty and the degree to which the item discriminates among examinees. For the teacher knowledge data, the empirical reliability of the item response theory estimates was .81.

Measuring teacher practice: The teacher practice observation protocol

After an extensive review of existing measures of teacher practice, as well as consultation with Technical Working Group members, the study team identified the Sheltered Instruction Observation Protocol (SIOP) (Echevarria, Vogt, and Short 2007) as the best instrument for measuring impacts on teacher practice with a focus on English language learner students. Although the instrument was originally designed as a tool for coaching and implementing the SIOP model, its developers claim that it can be used as an observation measure for evaluating teachers (Echevarria and Short 2004). Developers' research on the SIOP has identified reliable subscales (ranging from .92 to .95 for the three main dimensions) and shown that the protocol can distinguish between sheltered and nonsheltered instruction (Guarino et al. 2001).[123]

Modifications to the Sheltered Instruction Observation Protocol

The SIOP consists of 30 items, scored on a 5-point Likert scale from 0 to 4, with a "not applicable" option for 4 items. The 30 items cover 8 areas: preparation, building background, comprehensible input, strategies, interaction, application, effectiveness of lesson delivery, and lesson review/evaluation. In some studies, the developers grouped the eight areas into three dimensions or sections: preparation; instruction (includes building background, comprehensible input, strategies, interaction, application, lesson delivery); and evaluation/review (Echevarria and Short n.d.; Guarino et al. 2001).

To address the concern that the SIOP may not be effective for evaluating the reading comprehension and instructional techniques emphasized by Pacific CHILD, the study expanded the original SIOP with additional items. A pool of items was developed based on a review of the Pacific CHILD training materials and the Pacific CHILD observation protocol used by program staff during observations of teacher lessons. After the elimination of new items that overlapped with items already in the original SIOP protocol, eight additional items remained. Although these items were added to measure content covered in Pacific CHILD, they were designed to be general enough that well-trained teachers in both control and treatment groups could score well on them. With the additional items, the modified SIOP protocol contained 38 items

During the baseline year, an unmodified version of the 30-item SIOP was used to observe a convenience sample of classrooms in treatment and control group schools before the intervention. The expanded and modified SIOP protocol was used to observe teachers in the study schools toward the end of the first and second years of the intervention. Data collected toward the end of the second year of the intervention were used to measure the impact of Pacific CHILD on teacher knowledge.

---

[123] Sheltered instruction is an instructional approach in which teachers use specific strategies to make grade-level academic content comprehensible to English language learners while also promoting English language development (see, for example, Guarino et al. 2001).

Reliability of the teacher practice observation protocol

To ensure consistent use of the observation protocol in the field, classroom observers completed annual, week-long observation training before collecting data in the field. Observers were required to meet a minimum interrater reliability threshold of .80 for at least two out of three video ratings or live observations when paired with expert anchor raters in the field. They were required to meet the .80 level for both the original 30-item SIOP and the modified 38-item SIOP protocol. All trainings included live observations in classrooms. Interrater reliability was calculated using adjacent percent agreement, where scores within one point of each other were considered in agreement (for example, scores of 0 and 1 were considered in agreement, scores of 0 and 2 were considered in disagreement). For ratings of *not applicable*, observers were scored as in agreement only if both raters selected *not applicable*. Observers were also required to meet agreement thresholds for two of three video ratings during data collection to ensure that the observation protocol was used consistently throughout the data collection period.

Observation data used to assess the impact of Pacific CHILD on teacher practice were collected in the spring of the second year of the intervention. The average interrater reliability for observers during data collection in the spring of the second year was .91 for the 38-item SIOP protocol (.83–1.00), and .92 for the original 30 items (.84–1.00), with an average of 5.6 reliability checks per observer.

Outcome measures for teacher practice

Outcome measures for teacher practice were calculated as the average score, adjusted for not rated (*not applicable*) items. Preliminary analyses showed that one item, question 19 on first language use in the classroom, was not scored consistently across entities and did not contribute to the reliability of the SIOP. Because of the inconsistent scoring, this item was dropped from the calculation of the outcome measure for the benchmark impact analysis. The internal consistency of the modified SIOP was checked using Cronbach's alpha, a reliability estimate for responses with levels like Likert scale ratings.. The reliability of the observation data collected to measure teacher practice toward the end of the second year was .92. Dropping the inconsistently scored item (question 19) increased reliability to .92.

As an alternative scaling approach for the sensitivity analyses, the study used a Rasch partial credit rating model to generate ability measures for each teacher along the latent variable of teaching practice (Masters 1982). Partial credit Rasch models conceptualize each item as having its own unique rating scale structure. The items on the modified SIOP, although rated on a common scale of zero to four, have unique descriptors and different underlying scales.[124] Logits, the units of measurement in a partial credit Rasch model, are log-odds units that specify the probability, P_{nij}, that person n of ability measure B_n is observed in category j of a rating scale specific to item i of difficulty measure D_i as opposed to the probability, $P_{ni(j-1)}$, of being observed in category $(j - 1)$ of a rating scale with categories $j = 0, m$. The partial credit model

[124] For example, for item 1 on the SIOP, the scale ranges from *No language arts content objectives for students* (0) to *Language arts content objectives for current lesson implied for students* (2) to *Language arts content objectives for current lesson explicitly defined orally and in writing for students* (4); for item 25, the scale ranges from *Most students are paying attention and on task less than 50 percent of the period* (0) to *Most students are paying attention and on task approximately 70 percent of the period* (2) to *Most students are paying attention and on task 90 percent to 100 percent of the period* (4).

thus specifies the probability that a given teacher with a specific teaching ability will receive a score of 3 on item 1 versus the probability that the same teacher will receive a score of 2 on the same item. The software program Winsteps was used to calculate ability measures based on the Likert scores from the modified SIOP. The Rasch measures produced for the observation data collected at the end of the second year had a person reliability of .90.[125]

Limitations to the teacher outcome measures

There are several limitations to the teacher outcome measures. The teacher knowledge assessment was developed for the study and has not been used with other teacher samples or tested for construct validity against established measures of teacher knowledge. The 40-item assessment is also shorter than commercial assessments of teacher pedagogical knowledge,[126] which has implications for content coverage and the reliability of the resulting measure.[127] Although the assessment includes application questions, many of the questions are definitional and may not directly test the ability to apply knowledge of pedagogical concepts to classroom settings or scenarios.

As a measure of teacher practice, the SIOP also has limitations. The original SIOP was designed to provide educators with a protocol for assessing the quality of sheltered instruction (a teaching approach that integrates content and language instruction to support English language learners), not as an overall measure of teacher quality. No studies have validated the SIOP against other observational measures of teacher quality. Moreover, the additional items developed to measure components of Pacific CHILD have not been tested with other teacher populations.

---

[125] Person reliability in a Rasch model is analogous to test reliability in a classical testing theory framework. It reports how reproducible the order of person measures is for a given sample of people for a given set of items.

[126] For example, the computer-based Praxis I *Content and Structure* consists of 136 multiple choice questions; the Praxis II *English to Speakers of Other Languages* exam consists of 120 multiple choice questions.

[127] Test reliability is affected by the range and difficulty of items, the spread in test-taker scores, and the length of the assessment. For example, to attain a reliability of .80 for the impact sample, the teacher knowledge assessment would need an additional 4 items of similar quality to those in the assessment (for a total of 44 items).

Appendix E: Student outcome measures

This appendix provides additional information on the standardized assessments used to measure the impact of Pacific CHILD on student achievement in reading comprehension.

Student achievement data

The study used national, norm-referenced tests as the primary student outcome data. During the study period, the American Samoa Department of Education and the CNMI Public School System administered the Stanford 10 Achievement Test (SAT 10), and the Hawai‘i Department of Education administered the TerraNova, 2nd Edition, as part of their annual assessment of student achievement. American Samoa administered the SAT 10 to grade 4 students in the baseline, first, and second years. During the second year, it also administered the SAT 10 to grade 5 students. Throughout the study period, the CNMI administered the SAT 10 to grade 5 students. In Hawai‘i , students in grades 4 and 5 completed the TerraNova in the baseline, first, and second years. The tests administered to grade 5 students during March and April of the second year were used to measure the impact of Pacific CHILD.

During the study period, Hawai‘i also conducted its annual statewide standards-based assessment, the Hawai‘i State Assessment, to comply with the No Child Left Behind Act of 2001. American Samoa and the CNMI continued to develop and revise their content standards-based assessments and began implementing testing for selected grades during the study period. Because of uneven coverage of grades, the newness of the tests, and the lack of information on the content covered and the psychometric properties of the tests, none of the standards-based assessments were included as student outcome measures in the impact study.

Properties of standardized tests used to measure student reading comprehension

The SAT 10 and TerraNova are nationally administered tests that have been normed on national samples and vertically scaled across grades. Both assessments include a reading component with two subtests: reading comprehension and vocabulary. The reading comprehension subtest was selected as the outcome measure for the impact of Pacific CHILD on student achievement in reading. The length, average difficulty, and reliability estimates for the reading comprehension subtests on the TerraNova and SAT 10 for grade 5 are summarized in table E.1. The reading comprehension subtest from the SAT 10 is longer (54 items) and has higher reliability (.92), as measured using the KR-20 coefficient, than the reading comprehension subtest from the TerraNova (32 items and KR-20 of .89).

Table E.1 Length, average difficulty, and reliability of assessments used to measure reading comprehension in grade 5 students, by entity

Entity	Assessment	Length (number of items)	Mean *p*-value	KR-20
American Samoa	Stanford 10 Achievement Test	54	.58	.92
Commonwealth of the Northern Mariana Islands	Stanford 10 Achievement Test	54	.61	.92
Hawai‘i	TerraNova	32	.59	.88

Note: The mean p-value represents the average percentage of test-takers who correctly answers questions on the test based on the norming sample.

Source: Pearson (2004); CTB-McGraw Hill (2003).

Combining student achievement data

Two alternatives were explored to adjust for differences between the SAT 10 and the TerraNova in order to create a common metric for pooled analysis across the three entities: linking to create comparable scale scores and standardizing to *z*-scores within each entity subgroup. After considering both options, the study selected the linking method as the primary approach, with the *z*-score method used for the sensitivity analysis.

To link the SAT 10 and TerraNova scores and create a comparable scale score metric, the study converted reading comprehension scores from the TerraNova to estimated SAT 10 scores for Hawai‘i students using an equipercentile approach (Kolen and Brennan 2004). This approach links scores on one test to scores on another test with the same percentile ranks using a curve (nonlinear transformation) to capture potential variation between the two tests in difficulty along the score scale. Percentile-to-scale-score conversions from published norming tables, provided by Pearson and CTB-McGraw Hill, were used for the conversion. For example, a grade 5 student who scored 657 (50th percentile) on the TerraNova would receive a corresponding 50th percentile score of 643 on the Stanford 10. In the *z*-score approach, student scores were standardized to have a mean of 0 and a variance of 1 within each entity subgroup.

The main advantage of retaining scale scores as an outcome measure for this study was that it preserved the variance across entities. The *z*-score method requires standardizing and recentering scores within each subgroup to have a mean of 0 and a variance of 1. This recentering treats the average student within each group as equal (that is, both the average grade 5 student in Hawai‘i and the average grade 5 student in American Samoa receive equivalent *z*-scores of 0). Given the range of student ability across entities, using *z*-scores reduces the overall variability in the full student sample and eliminates between-entity variation in scores. For example, using linked scale scores, the same average grade 5 student would have a score of 597 in American Samoa and 638 in Hawai‘i (table E.2)Descriptive statistics for linked scale scores (SAT 10 and SAT 10 equivalents) and the *z*-scores are provided in table E.2.

Table E.2 Scale scores, linked scale scores, and *z*-scores for measures of student achievement in reading comprehension

	Linked scale score (SAT 10 or SAT 10–equivalent)				*z*-score			
Entity	**Mean**	**Standard deviation**	**Minimum**	**Maximum**	**Mean**	**Standard deviation**	**Minimum**	**Maximum**
American Samoa	597	21	552	663	0.00	1.00	–2.17	3.19
Commonwealth of the Northern Mariana Islands	631	30	552	722	0.00	1.00	–2.66	3.08
Hawai‘i	638	40	556	734	0.00	1.00	–2.77	2.16

SAT 10 is the Stanford 10 Achievement Test.

Note: Figures are for data collected toward end of second year of intervention.

Source: Authors' calculations based on the data collected for the study.

Limitations to the student outcome measure

The main limitation of the outcome measure for student achievement in reading comprehension is that it is a pooled outcome measure across two different tests, the SAT 10 and the TerraNova. No studies have explored the extent to which the two tests measure the same or similar content in reading comprehension. Although both tests are normed on national samples, no studies have compared the difficulty levels of the two tests or conducted concordance studies to examine the relationship between scores on the two tests.

Appendix F: Adaptations to design and context of implementation

This appendix documents primary adaptations made to Pacific CHILD during the implementation period and describes the contextual factors related to program implementation. It supplements the interpretation of findings presented in chapter 4.

Adaptations to the intervention

Chapter 4 compares the intervention as designed with the actual amount of exposure to the intervention by teachers at treatment group schools in the impact sample and the actual delivery of program activities by program staff in the field. The focus of this appendix is on adaptations to the original design of the Pacific CHILD professional development program.

Adaptations to the prescribed level of participation

In addition to the annual institute and the mini-institutes, each teacher participating in Pacific CHILD was expected to receive four hours of support from program staff each month. During the course of the two-year intervention, 12 teachers were placed on a modified treatment plan.[128] Program staff provided modified treatment—which excluded classroom observations and, in some cases, structured learning teams—in an effort to retain teachers who might otherwise have dropped out of the intervention.[129]

Adaptations to the program structure

During the two-year intervention, REL Pacific made the following primary adaptations to the program structure:

- Pre- and postconferences for classroom observations and lesson demonstrations in two of the three entities were conducted in person, by phone, and by email rather than in person, as specified in the professional development manual. Across all entities, preconferences were not consistently delivered before the lesson demonstrations.
- During the first year, program developers reduced the length of classroom observations from 75 to 60 minutes, based on the amount of instructional time available during the English language arts period. The delivery of year-round classroom observations and lesson demonstrations was not limited to English language arts instruction but included subjects such as social science or science.

[128] Six teachers received a modified plan during the first year of implementation. Twelve teachers received a modified plan during the second year.

[129] Most teachers on modified plans reduced their year-round instructional support activities to demonstration lessons. The modified plan varied based on individual teacher needs and availability to participate throughout the school year. Six teachers limited their participation to demonstration lessons for one or more months in Year 1. Ten teachers limited their participation to demonstration lessons for one or more months in Year 2.

- The project advisor did not attend every institute, as specified in the professional development manual. The project advisor attended mini-institutes 2 and 5 in person, instead of via technology, as specified in the professional development manual. School principals did not consistently attend portions of the mini-institutes, as specified in the professional development manual.
- In schools in areas without locally based program staff, mini-institute school-based activities were combined and delivered in one day rather than staggered across two consecutive days, as specified in the professional development manual. At these schools, year-round instructional support activities (lesson demonstrations, classroom observations, and structured learning teams) also frequently occurred on the same day. At some schools in areas with locally based program staff, mini-institute school-based days were extended across a two- to three-week period, in some cases, delaying the implementation of the regular year-round activities. Mini-institute school-based lesson demonstrations were held before the full-day mini-institute workshop at nine treatment group schools in two entities. In schools that did not provide release time for teachers to observe lesson demonstrations in other teachers' classrooms, program staff delivered separate demonstrations in each participating teacher's classroom.
- Mini-institutes were not always delivered as designed. In some cases, the full day portion of the mini-institutes was held on weekends, split into two days, or conducted separately for each grade.
- After the first year of implementation, program staff supplemented structured learning teams during the annual institute and mini-institutes with the use of cooperative learning groups, collaborative teacher groups that consisted of teachers from across rather than within schools.
- Annual institutes were designed to provide five days of workshop-style professional development and five days of onsite practice opportunities with students at a local elementary school. Each day of the second week of the annual institute was designed to include a demonstration lesson by program staff, followed immediately by lessons delivered by participating teachers. During the second year, program staff modified the design of the annual institute in response to teachers' requests for more time to incorporate what they learned from the staff demonstration in preparing their own practice lessons. The design of the annual institute was modified to include two lesson demonstrations on Day 6, which provided teachers with more time to modify and plan their own classroom lessons after observing the demonstration by program staff. As a result, Week 2 included the following sequence:
    - Day 6: Two lesson demonstrations.
    - Days 7–9: One lesson demonstration followed by teacher lessons.
    - Day 10: Two teacher lessons or an extended teacher lesson.

Adaptations to program content

Program staff made adaptations to two of the six Pacific CHILD components after the first year of implementation in two entities (during the 2007/08 school year). They also modified the component delivery schedule (table F.1).

The first adaptation was the replacement of the focus on sequence with a focus on text features. The second was a change in the question-generation component. Originally, the focus of the question-generation component was on developing students' use of literal, inferential, and evaluative questions. During the first year of implementation, program staff modified the focus of this component to developing students' use of questions to find the main ideas in and answer questions about the text. As a result of these changes and the staggered implementation across entities, the focus on sequence and the focus on question generation were not implemented in one entity during the two-year implementation period (2008/09 and 2009/10 school years). However, all six primary Pacific CHILD component categories (vocabulary, text structure, question generation, differentiated instruction, cognitively rich environment, and interactive tasks) were delivered in the three study entities during the implementation period.

Table F.1 Adaptations to the component delivery schedule

		As implemented	
Component	**As designed**	**2007–09 (across two entities)**	**2008–10 (across one entity)**
Vocabulary: Word knowledge	Terms 1, 3, 4, and 6	Terms 1, 4, and 6	Terms 1, 4, and 6
Vocabulary: Word parts	Terms 4, 5, and 6	Terms 3, 4,[a] and 6	Terms 3, 4,[a] and 6
Text structure: Sequence[b]	Terms 1 and 5	Terms 1	—
Text structure: Text features[b]	—	Terms 5	Terms 1 and 5
Text structure: Compare and contrast	Terms 2 and 4	No change	No change
Text structure: Cause and effect	Terms 3 and 6	No change	No change
Question generation	Terms 2, 5, and 6	Terms 2 and 5	Terms 2 and 5
Differentiated instruction	Terms 3 and 6	Terms 3 and 4	Terms 2 and 4
Cognitively rich environment	Terms 2 and 5	Terms 2 and 6	Terms 3 and 6
Interactive tasks	Terms 1 and 4	Terms 1 and 5	Terms 1 and 5

— is not available.

a. In Term 4, word parts were either delivered during the full-day mini-institute workshop or, in some cases, were incorporated into the school-based days of mini-institute 4.

b. Text features replaced sequence during the first year of implementation in two entities during the 2007/08 school year.

Source: Authors' analysis of professional development manual and professional development observation field notes.

Context of implementation

This section describes contextual factors related to program delivery and teacher participation.[130] It provides a context for interpreting the findings presented in the previous sections regarding the extent to which Pacific CHILD was implemented as designed.

The findings in this section are based on systematic reviews of data obtained from focus groups with teachers at treatment group schools, interviews with program staff, and observations of intervention activities. Appendix K provides a description of the qualitative data analysis used in this section. Qualitative data from these sources are not intended to be quantifiable, to be generalizable to the entire treatment group, to represent the views of all program staff, or to determine causal inferences or to draw conclusions about the impact or exploratory analysis presented in this report. Instead, these data are intended to describe the context within which Pacific CHILD was implemented. Appendix K provides an overview of the qualitative analysis presented here.

Analysis of the focus group, interview, and observation data identified seven primary barriers to participation and implementation: miscommunication regarding expected participation; lack of alignment with district- or school-level mandates; the use of expository text; teacher time constraints; lack of substitute teachers; program staffing; and furloughs in Hawai‘i. These barriers to program fidelity are described below.

Miscommunication regarding expected participation

Teachers from five schools in one entity found that Pacific CHILD required a greater time commitment than they initially expected. Although recruitment presentations included a description of the Pacific CHILD activities provided during the two-year intervention, teachers from five schools thought that Pacific CHILD was a one-year intervention. Four teachers left the program after they learned during the first annual institute that the program lasted two years.

Lack of alignment with district- or school-level mandates

The scope and sequence of the Pacific CHILD component delivery did not always align with district grade-level English language arts instructional pacing calendars (see table 4.1 in chapter 4 for the Pacific CHILD component calendar). Although Pacific CHILD was developed to align with grade 4 and grade 5 English language arts standards, teachers reported challenges incorporating specific Pacific CHILD content into their instruction while following district-mandated instructional schedules. For example, the cause and effect component of Pacific CHILD was designed to be delivered in the spring. However, in some schools, pacing guides directed teachers to deliver cause and effect instruction in the fall.[131] Teachers from the six schools undergoing restructuring under the No Child Left Behind Act of 2001 during the study noted that they found it particularly difficult to integrate lessons based on Pacific CHILD content if the lessons were not aligned with the instructional frameworks they were expected to deliver under school-wide reform programs.

---

[130] The data examined in this section include the three treatment group schools that were excluded from the impact analysis.

[131] Pacing guides provide a general guide for teachers in the implementation and delivery of instruction.

Use of expository text

Participating teachers and program staff reported that Pacific CHILD's emphasis on expository text was a barrier to implementation in classrooms that used primarily narrative texts in their English language arts curricula. Pacific CHILD was designed to supplement rather than replace existing curriculum. However, teachers reported that it was a challenge to locate grade-appropriate expository texts that were both aligned with grade-level standards and could be used for Pacific CHILD lessons.

Teacher time constraints

Teachers noted that personal obligations (family vacations, caring for children or parents, attending continuing education classes) and school-related obligations outside of instructional time (parent–teacher conferences, lunch or recess yard supervision, committee work) conflicted with school-based Pacific CHILD activities. In addition, treatment group schools in one entity had a contractual teacher union agreement that placed restrictions on teacher participation in school-related activities beyond a seven-hour work day. Teachers in these schools did not consistently attend structured learning teams after school hours.[132] Program staff also reported that it was challenging to coordinate the schedules of participating teachers given the restrictions on their availability.

Lack of substitute teachers

Teachers in treatment group schools were not always able to leave their classrooms during instructional time, because of lack of substitute teachers. To address this problem, in the first year of implementation, the study held the full day of the mini-institute on Saturday at some sites. During the second year, the full day of the mini-institute was delivered during after-school hours on two consecutive afternoons in one entity. In another entity, it was separated by grade and delivered across two full school days.[133]

Program staffing

Pacific CHILD was designed to use local staff to deliver the intervention to treatment group schools. However, local program staff were not available to all treatment group schools. Five treatment group schools in the impact sample were not served by local staff. In these schools, program staff from other islands had to travel to the schools to deliver school-based support to teachers. At these schools, school-based activities were often delivered during a one-day visit rather than staggered across two or more days. Teachers without access to local program staff often rearranged their scheduled English language arts block to accommodate program staff who traveled from other islands for lesson demonstrations and classroom observations.

A general shortage of program staff to serve all participants was also a challenge to program fidelity, particularly in areas with geographically distant schools. In areas with a high teacher-to-

---

[132] Some of these teachers attended meetings of structured learning teams during lunch instead.

[133] Program staff in this entity secured release time for teachers for mini institutes 4-6. However, only one grade level could be released each day. As a result, two separate mini-institutes, based on grade level, were provided to teachers during the second year.

program staff ratio, mini-institute days often extended beyond the designed three days because of lack of sufficient program staff to deliver these activities within the intended time frame.[134] Extended mini-institutes delayed or reduced the delivery of other year-round support activities. Teachers and program staff reported that the shortage of staff was particularly challenging at schools that did not have the resources to release teachers from class during school hours. In these schools, staff had to deliver demonstration lessons in each participating teacher's classroom each month rather than provide one monthly demonstration lesson attended by all participating teachers.

Furloughs in Hawai'i

During the second year of implementation in Hawai'i, instructional days were reduced by 17 days in response to the state's budget deficit (Hawai'i Department of Education n.d.). The cutback in instructional days reduced the number of days program staff could deliver mini-institutes and year-round support activities. In addition, teachers who were required to follow instructional pacing guides had less time to prepare for Pacific CHILD activities, particularly classroom observations. Teachers reported that the loss of instructional days made it difficult for them to meet both district- and school-level English language arts requirements while planning for and incorporating Pacific CHILD lessons.

[134] During the intervention period, the teacher-to-program staff ratio ranged from 3:1 to 18:1, with an average of 10 teachers per program staff person.

Appendix G: Exposure to Pacific CHILD by students in the impact sample

In order to contextualize the estimated effects of Pacific CHILD on students, it is important to understand how much of the intervention students in the impact sample received. This appendix describes issues related to student exposure to the intervention and the limitations of the study in measuring that exposure.[135]

Pacific CHILD was designed and implemented as a school-level intervention targeting teachers rather than students. As such, the design of Pacific CHILD did not specify the expected form or level of the intervention for students. Conceptualizing students' exposure to Pacific CHILD without design-based definitions for what that exposure entails is a complex task, for two reasons. First, students could have been exposed to the intervention both directly (through program staff who regularly visited the classrooms of actively participating teachers) and indirectly (through their classroom teachers who actively participated in the program). Students could also have been exposed indirectly through their interactions with other teachers and students at their schools who were exposed to the intervention. Second, year-to-year reassignment of classrooms could affect individual students' direct and indirect access to the intervention over the two-year intervention period, as not all classroom teachers participated in the intervention. (For example, a student may have had an actively participating teacher one year and a nonparticipating teacher another year.) Measuring exposure to the intervention by students thus requires weighting and aggregating students' cumulative exposure through their classroom teachers (and other teachers in the school) across years.[136]

One practical approach is to assess students' exposure to the intervention only in terms of the experience of their assigned classroom teachers. This approach is consistent with a theoretical model in which the main conceptual link between the intervention and student outcomes is through their classroom teachers. Measuring exposure in this way was not possible, because the study lacked data with which to reliably link individual students to their classroom teachers.

An alternative approach is to approximate students' exposure in terms of their schools' exposure to the intervention. This approach is consistent with the primary research question, which focused on the impacts on students at schools offered the intervention. One advantage of this approach is that it does not require defining how to measure each student's exposure to the intervention. School-level exposure could be summarized, for example, by the total amount of professional development services received by all teachers who participated in the intervention at a given school over the two-year implementation period. The limitation of this approach is that it does not reflect how each individual student experienced the intervention. Instead, it describes the context in which most of the students in the impact sample were exposed to the

[135] As explained in chapter 2, the student impact sample consisted of grade 5 students who were at study schools at the end of the two-year implementation. In measuring exposure, the discussion here concerns the treatment group in the student impact sample.

[136] For example, a teacher could have participated in the intervention in both years, during only the first, during only the second year, or in neither year. Analyzing exposure using this approach would thus require determining how to assess student exposure in each of these cases.

intervention.[137] School-level exposure to the intervention in treatment group schools can be described by the following set of indicators:

- Total professional development hours provided to all 23 treatment group schools over two years: 14,222 hours (2,032 school days).
- Total professional development hours provided per treatment group school over two years: 585 hours (84 school days).
- The average number of hours of professional development provided per program-eligible teacher per treatment group school over two years: 117 hours (17 school days), with a standard deviation of 69 hours. The average number of hours per eligible teacher ranged from 23 to 264 in those schools with participating teachers.

The professional development hours include the total hours of all institutes and year-round support activities provided to all program-eligible teachers at treatment group schools over the two years. These numbers provide a context for understanding the impacts on students.

[137] As explained in chapter 2, the student impact sample consisted of grade 5 students who were at the study schools at the end of the two years of implementation. About 75 percent of these students were estimated to have been enrolled continuously at the same study schools for two years.

Appendix H: Exposure to Pacific CHILD by participating teachers in the impact sample

This appendix provides information on the exposure to Pacific CHILD of treatment impact teachers who participated in intervention activities during the two year intervention. Participation was defined as attendance of more than zero hours.

Of the 118 teachers at treatment group schools, 68 participated in at least one activity during the two years. Activity included attendance at the annual institute (table H.1) and mini-institutes (table H.2) as well as receipt of year-round support (table H.3). Exposure to Pacific CHILD from all three sources is summarized in table H.4.

Table H.1 Prescribed and actual attendance at Pacific CHILD annual institute by participating teachers at treatment group schools in impact sample

		Number of attendees by days of attendance					
Annual institutes	**Number of attendees**	**1–4 days**	**5–7 days**	**8–10 days**	**Prescribed days of attendance**	**Average days attended**	**Percentage of prescribed intervention received[a]**
Annual instiute1	62	9	8	45	10	8.0	80.0
Annual institute 2	49	8	5	36	10	8.0	80.3
Average	56	na	na	na	10	8.0	80.2

na is not applicable.

a. Average days reported are rounded. Percentages reported are based on unrounded average days.

Source: Authors' analysis of annual institute teacher sign-in sheets.

Table H.2 Prescribed and actual attendance at Pacific CHILD mini-institutes by participating teachers at treatment group schools in impact sample

Mini-institute	Number of attendees	Number of attendees, by days of attendance: 1 day or less	Number of attendees, by days of attendance: More than 1 day	Prescribed days of attendance	Average days attended	Percentage of prescribed intervention received[a]
Mini-institute 1	55	5	50	2	1.9	94.1
Mini-institute 2	50	8	42	2	1.8	90.8
Mini-institute 3	48	9	39	2	1.8	89.6
Average mini-institutes 1–3	51	na	na	2	1.8	91.5
Mini-institute 4	53	9	44	2	1.8	91.5
Mini-institute 5	53	9	44	2	1.8	90.6
Mini-institute 6	53	15	38	2	1.7	85.4
Average mini-institutes 4–6	53	na	na	2	1.8	89.2
Average mini-institutes 1–6	52	na	na	2	1.8	90.3

na is not applicable.

a. Average days reported are rounded. Percentages reported are based on unrounded average days.

Source: Authors' analysis of mini-institute teacher sign-in sheets.

Table H.3 Prescribed and actual year-round Pacific CHILD support per month for participating teachers at treatment group schools in impact sample

Year	Number of teachers who attended at least one year-round activity	Number of teachers, by hours of attendance		Prescribed hours	Average actual hours	Percentage of prescribed intervention received[a]
		Less than 2 hours	2 or more hours			
First year	56	23	33	4	2.3	56.4
Second year	58	28	30	4	1.9	48.4
First and second year	57	Na	na	4	2.1	52.4

na is not applicable.

a. Average days reported are rounded. Percentages reported are based on unrounded average days.

Note: Attendance at the twice-monthly teacher-only structured learning team meetings was not recorded in the participation data and therefore not included in the prescribed or average year-round support activities. The average number of hours of year-round support per month was determined by dividing the total number of year-round support activity hours provided per year by the nine months of the school calendar year.

Source: Authors' analysis of year-round support monthly activity logs.

Table H.4 Prescribed and actual exposure to Pacific CHILD by participating teachers at treatment group schools in impact sample

Year	Prescribed exposure (days)	Average exposure of teachers at treatment group schools (days)	Prescribed intervention received (percent)[a]
First year	21.1	16.8 (117.4 hours)	79.5
Second year	21.1	14.6 (102.2 hours)	69.2
First and second year	42.2 (295 hours)	31.4 days (219.6 hours)	74.3

a. Average days reported are rounded. Percentages reported are based on unrounded average days.

Note: Table shows exposure to Pacific CHILD from annual institutes, mini-institutes, and year-round support. Attendance at the twice-monthly teacher-only structured learning team meetings was not recorded in the participation data and therefore not included in figures reported here. Seven-hour days were used to convert activities measured in hours to days to calculate total year-round support received per year and total hours of prescribed intervention. Average exposure includes the 54 teachers who attended intervention activities during both the first and second year of the intervention.

Source: Authors' analysis of annual institute and mini-institute teacher sign-in sheets and program staff monthly activity logs.

Appendix I: Exposure to Pacific CHILD by teachers in impact sample who were present at treatment group schools during entire intervention period

This appendix provides information on the exposure to Pacific CHILD by the 97 participating treatment group teachers in the impact sample who were present at treatment group schools during the entire two-year intervention period. It documents their participation in the annual institute (table I.1) and mini-institutes (table I.2), as well as their receipt of year-round support (table I.3). Exposure to Pacific CHILD from all three sources is summarized in table I.4.

Table I.1 Prescribed and actual attendance at Pacific CHILD annual institute by teachers in impact sample who were present at treatment group schools during entire intervention period

Annual institute	Number of attendees	Number of attendees by days of attendance				Prescribed days of attendance	Average days attended	Percentage of prescribed intervention received[a]
		0 days	1–4 days	5–7 days	8–10 days			
Annual institute 1	60	37	9	7	44	10	4.9	49.3
Annual institute 2	43	54	4	5	34	10	3.8	37.7
Average	52	na	na	na	na	10	4.3	43.5

na is not applicable.

a. Average days reported are rounded. Percentages reported are based on unrounded average days.

Source: Authors' analysis of annual institute teacher sign-in sheets.

Table I.2 Prescribed and actual attendance at Pacific CHILD mini-institutes by teachers in impact sample who were present at treatment group schools during entire intervention period

Mini-institute	Number of attendees	Number of attendees by days of attendance			Prescribed days of attendance	Average days attended	Percentage of prescribed intervention received[a]
		0 days	1 day or less	More than 1 day			
Mini-institute 1	52	45	5	47	2	1.0	50.3
Mini-institute 2	48	49	8	40	2	0.9	44.7
Mini-institute 3	46	51	9	37	2	0.8	42.3
Average of mini-institutes 1–3	49	na	na	na	2	0.9	45.7
Mini-institute 4	48	49	8	40	2	0.9	45.4
Mini-institute 5	49	48	9	40	2	0.9	45.4
Mini-institute 6	47	50	12	35	2	0.8	42.0
Average of mini-institutes 4–6	48	na	na	na	2	0.9	44.2
Average of mini-institutes 1–6	48	na	na	na	2	0.9	45.0

na is not applicable.

a. Average days reported are rounded. Percentages reported are based on unrounded average days.

Source: Authors' analysis of mini-institute teacher sign-in sheets.

Table I.3 Prescribed and actual year-round Pacific CHILD support per month of teachers in impact sample present at treatment group schools during entire intervention period

Year	Number of teachers who attended at least one year-round activity	Number of attendees, by hours of attendance			Prescribed hours	Average hours	Percentage of prescribed intervention received[a]
		0 hours	Less than 2 hours	2 or more hours			
Year 1	53	44	20	33	4	1.29	32.3
Year 2	51	46	23	28	4	1.12	28.0
Years 1 and 2	104	90	43	61	4	1.21	30.3

a. Average days reported are rounded. Percentages reported are based on unrounded average days

Note: Attendance at the twice-monthly teacher-only structured learning team meetings was not recorded in the participation data and therefore not included in the prescribed or average year-round support activities. The average number of hours of year-round support per month was determined by dividing the total number of year-round support activity hours provided per year by the nine months of the school calendar year.

Source: Authors' analysis of year-round support monthly activity logs.

Table I.4 Prescribed and actual exposure to Pacific CHILD teachers in impact sample who were present at treatment group schools during entire intervention period

Year	Prescribed exposure (days)	Average exposure of teachers at treatment group schools (days)	Percentage of intervention received[a]
Year 1	21.1	9.3	44.2
Year 2	21.1	7.9	37.3
Years 1 and 2	42.2 (295 hours)	17.2 (120 hours)	40.7

a. Average days reported are rounded. Percentages reported are based on unrounded average days.

Note: Table shows exposure to Pacific CHILD from annual institutes, mini-institutes, and year-round support. Attendance at the twice-monthly teacher-only structured learning team meetings was not recorded in the participation data and therefore not included in figures reported here. Seven-hour days were used to convert activities measured in hours to days to calculate total year-round support received per year and total hours of prescribed intervention.

Source: Authors' analysis of annual institute and mini-institute teacher sign-in sheets and program staff monthly activity logs.

Appendix J: Teachers' exposure to Pacific CHILD by entity

This appendix provides information on the exposure to Pacific CHILD by teachers at treatment group schools in the impact sample by entity (table J.1). The information is provided to document the variation in exposure to the intervention by teachers across the three entities.

Table J.1 Total exposure to Pacific CHILD by teachers at treatment group schools in impact sample, by entity

Entity	All teachers	Did not participate in any activity	Participated in at least one activity
American Samoa			
Number of teachers	10	0	10
Average days exposure per teacher	27.2	0	27.2
Commonwealth of the Northern Mariana Islands			
Number of teachers	33	14	19
Average days exposure per teacher	16.6	0	28.8
Hawai'i			
Number of teachers	75	36	39
Average days exposure per teacher	12.7	0	24.4
All entities			
Number of teachers	118	50	68
Average days exposure per teacher	15.0	0	26.0

Source: Authors' analysis of annual institute and mini-institute teacher sign-in sheets and program staff year-round support monthly activity logs.

Appendix K: Qualitative methods used to analyze adaptations to design and context of implementation

This appendix provides an overview of the qualitative analysis presented in appendix F, which examines differences between actual and planned implementation of Pacific CHILD. The analysis was guided by the following questions:

- Was Pacific CHILD implemented as designed?
- What factors may have affected program fidelity and teacher participation?

The first question examines how much professional development was received by teachers, whether program staff implemented activities according to the program design, whether teachers received the intervention content as planned, and how the design was modified during the first and second year of implementation.

The qualitative data analysis was originally intended to be an integral part of the mixed research methods used in this study.[138] The implementation study questions listed above were to have been explored using all available data sources, including quantitative data. The purpose of this appendix is to focus on the description of the qualitative data and analysis presented in appendix F, as guided by the above research objectives.

Three indicators were developed to determine if the Pacific CHILD program was delivered with fidelity to the professional development manual developed by REL Pacific (table K.1):

- Participation: Amount of program intervention received by teachers.
- Delivery and structure of intervention activities: Extent to which activity structure was delivered as designed by program staff.
- Content: Extent to which teachers received the six core components of Pacific CHILD. .

[138] The study of Pacific CHILD was originally designed as a comprehensive mixed-method evaluation project. This report focuses on the impact study component of the evaluation.

Table K.1 Indicators for determining fidelity of implementation to Pacific CHILD design

Indicator	Description
Participation	Measures extent to which teachers participated in intervention activities as designed. Indicator is based on teacher participation in the following activities: • Two annual 2-week summer institutes • Six 3-day mini-institutes • Once-monthly demonstrations by program staff • Twice-monthly classroom observations • Weekly structured learning team meetings
Delivery and structure of activities	Measures extent to which Pacific CHILD activity structure was implemented by program staff as designed. Indicator includes duration and frequency of the following activities: • Two annual 2-week summer institutes • Six 3-day mini-institutes • Once-monthly demonstrations by program staff • Twice-monthly classroom observations • Weekly structured learning team meetings
Content	Measures extent to which teachers received the six core intervention components: • Vocabulary • Text structure • Question generation • Differentiated instruction • Cognitively rich environment • Interactive tasks

Source: Authors' summary of REL Pacific teacher manual (Pacific Resources for Education and Learning 2007).

Data sources and preparation

The following sections describe the sources of data used to answer the implementation research questions.

Interviews with program staff

Telephone interviews were conducted with program developers, project advisors, and trainers to gather information on the study conditions and to determine the extent to which the intervention had been implemented with fidelity. Interviews were conducted with Pacific CHILD staff using a semistructured interview protocol to gather information on the background and training of the trainers, delivery of the intervention, teacher- and school-level implementation successes, and barriers encountered during implementation. Telephone interviews were conducted from May to October during the first and second years of the intervention.

Eighteen of 29 interviews with program staff were audiotaped and transcribed. Program staff interviews that were not transcribed were documented with detailed notes. A lead researcher reviewed each program staff interview to ensure that all interview questions had been comprehensively addressed and to determine whether additional follow-up was needed.

Interview transcripts and notes were saved as Word files and imported into the NVivo qualitative software program for analysis.

Focus groups and group interviews with teachers

Focus groups were conducted with teachers in treatment group schools to examine the extent to which they received the intervention as intended. The semistructured focus group protocol contained open-ended questions designed to elicit information about teachers' perceptions of the program and the training they received, satisfaction with and commitment to the intervention, and issues they may have encountered implementing Pacific CHILD. These focus groups were conducted between January and May during the first year of the intervention and between February and May during the second year of the intervention. All teachers who participated in the intervention were invited to participate in the focus groups.

Twenty of 28 focus groups were audiotaped and transcribed. Focus groups that were not transcribed were documented with detailed notes. Each focus group facilitator read the transcripts and condensed the raw transcript or detailed notes into a summary format. A lead researcher reviewed the transcripts and notes to ensure that the summaries highlighted the adaptations to the design and included all of the key themes that emerged from the focus group transcripts. Focus group transcripts, notes, and summaries were saved as Word files and imported into the NVivo qualitative software program for analysis.

Observations of Pacific CHILD professional development activities

Activity observation protocols and observation data collection forms were designed to capture the extent to which the intervention was implemented with fidelity. The professional development observation protocols required observers to compare the designed activities, as outlined in the Pacific CHILD professional development manual, with the actual activities observed. The protocols also required observers to summarize the fidelity of program delivery for each professional development activity. The professional development observation protocol fidelity criteria were based on the Pacific CHILD professional development manual and interviews with program staff. Observations of Pacific CHILD activities were conducted throughout the calendar year during the first and second years of the intervention. The field research team observed the annual institute and mini-institutes at each location and a sample of the structured learning team meetings at each treatment group school.[139]

Site visit reports included a chart for comparing the extent to which specified components were delivered as prescribed and the amount of program content participants received. The observation summaries also documented specific instances of adaptation of the model and the frequency with which adaptations occurred. A lead researcher reviewed the professional development site visit reports and implementation data, cross-checking the detailed notes against the summary and tables for accuracy.

---

[139] Mini-institute observations included a sample of school-based classroom observations and lesson demonstrations. Fewer than three mini-institutes were not observed in American Samoa during the first year because of a schedule changes.. Institutes were conducted at the entity level in American Samoa and the CNMI and at two to five locations in Hawai'i.

Training

All program staff interviews and focus groups were led by experienced senior staff. As part of their training, focus group facilitators, interviewers, and coders were required to read the Pacific CHILD professional development manual to become familiar with the original design and recognize deviations from implementation as designed. Interviewers and focus group facilitators were also trained on the use of the protocols and on questioning techniques to use during interviews and focus groups. Professional development activity observers received comprehensive training on the objectives of the observations, the use of nonparticipant observation techniques, and note-taking skills. They also received guidance on the detailed site visit reporting process, including tips on recognizing and reporting deviations from the original activity design.

After initial training, coders were asked to code a sample data file, which the lead qualitative researcher then reviewed. The coder and the lead researcher then discussed the codes used in the data file and reconciled any coding discrepancies. After each coder completed training, a senior researcher reviewed a sample of each data file for accuracy. The frequent and consistent use of queries in NVivo enabled the lead researcher to check the accuracy of coding throughout the coding process.

The coding process was led by a senior researcher who was familiar with and reviewed all of the qualitative data for this study. Senior research staff coded all interview and focus group summaries and transcripts. Senior and midlevel staff coded all professional development activity observations.

Qualitative analysis

The qualitative software package NVivo was used to facilitate the analysis of qualitative data collected from focus groups with teachers, interviews with program staff, and observations of intervention activities. The analysis process involved identifying and examining themes and patterns that emerged from the data. Both deductive and inductive approaches to data analysis were used to code qualitative data in NVivo and to systematically analyze the data. In the deductive approach, codes aligned with the research objectives were developed before data analysis. In the inductive approach, patterns and themes in the data led to the development of emergent codes (codes that correspond to themes or patterns that emerge or reoccur in the data). The following sections describe the development of preestablished and emergent codes using both deductive and inductive approaches.

Pre-established codes

Pre-established codes were developed based on the implementation research objectives . These codes included the three fidelity indicators of participation, delivery structure, and content (see table K.1) and codes regarding potential barriers to implementation of Pacific CHILD. Data that captured factors associated with teachers' ability to attend the prescribed number of intervention activities were coded for participation fidelity. To examine the delivery and structure of intervention activities, the study used pre-established codes (such as staffing constraints, adaptations, and so forth) to examine the extent to which intervention staff were able to deliver

the intervention activities as planned. Pre-established codes were also used to examine the extent to which teachers received intervention content. The content codes were applied to capture deviations from the core intervention content as outlined in the professional development manual. Pre-established codes also included the professional development activity type (annual institute, classroom observation, and so forth). Each code often included several subcodes, which further defined the code categories (see sample code list below in table K.2).

Table K.2 Sample of pre-established codes used to categorize Pacific CHILD annual institute and mini-institute observations

1 Training Format/Activity
- 1.1 Lecture
- 1.2 Small Group Work
- 1.3 Group Discussion
- 1.4 Demonstrations
- 1.5 Classroom Observations
- 1.6 Team-Teaching
- 1.7 Structured Learning Team
- 1.8 Individual Teaching
- 1.9 Lesson Planning
- 1.10 Evaluation
- 1.11 Taking Stock
- 1.12 Preconference: This node should be used in combination with Classroom Observations and Demonstrations to clarify which it is referring to.
- 1.13 Postconference: This node should be used in combination with Classroom Observations and Demonstrations to clarify which it is referring to.

2 Training Content
- 2.1 Vocabulary
  - 2.1.1 Word Knowledge
  - 2.1.2 Word Parts
- 2.2 Text Structure
  - 2.2.1 Sequence
  - 2.2.2 Compare and Contrast
  - 2.2.3 Cause and Effect
  - 2.2.4 Text Features
- 2.3 Question Generation
- 2.4 Differentiated Instruction
- 2.5 Cognitively Rich Environment
- 2.6 Interactive Tasks

Source: Authors.

The pre-established fidelity indicator and factors that may have affected program fidelity and teacher participation codes were not mutually exclusive. For example, school obligations reported by teachers that conflicted with after-school structured learning team meetings were

coded under “training format/activity” (specifically, the subcode “structured learning team””); “school obligations” (a predetermined potential barrier to implementation); “participation fidelity”“; and “challenges/barriers”.” Data coded to fidelity indicators were also cross-coded with factors that may have affected implementation of Pacific CHILD. Cross-coding in NVivo enabled the research team to query for “participation fidelity” and “challenges/barriers” to locate text about the challenges associated with teacher participation (fidelity indicator 1).

Emergent codes

Emergent codes, codes that correspond to themes or patterns that emerge or reoccur in the data, were also used to code the qualitative data. These codes were identified by examining and reexamining the interview, focus group, and professional development observation data. One code that emerged and reoccurred in the data was the use of expository text. The emphasis on expository text in classrooms that used primarily narrative texts in their English language arts curricula was an unanticipated barrier to implementation of Pacific CHILD that emerged from the coding process. All data were recoded after all emergent codes were identified to ensure that all documents were comprehensively coded with the full code list.

Attributes

Attributes were applied to all qualitative summaries and transcript documents uploaded into the NVivo software. An attribute is a characteristic assigned to a data file that allows for further classification and organization of the data. For example, documents were assigned attributes to allow classification according to study year (first year, second year); study entity (American Samoa, the CNMI, Hawai‘i); and data type (teacher focus group, program staff interview, annual institute week 1, and so on). Once coded with attributes, data were queried according to particular characteristics. For example, all instances of barriers to implementation in American Samoa (study entity) during the first year (study year) focus groups (data type) were queried to examine patterns within American Samoa during the first year focus groups.

Cross-verification of data through triangulation

Multiple data sources were used to triangulate and confirm the findings presented in interviews with program staff, focus groups with teachers, observations of Pacific CHILD activities, and attendance logs (table K.3). All qualitative data were analyzed for consistent patterns across multiple data sources. The use of multiple sources of information and multiple analysis methods ensured that the findings were verified and supported from different perspectives.

Table K.3 Sources of data for assessing fidelity of Pacific CHILD implementation

Measure of fidelity of implementation	Interviews with program staff	Teacher focus groups/group interviews	Observations of Pacific CHILD activities	Dosage and attendance logs
Teacher participation: How much professional development did teachers receive?				x
Activity structure: Were activities delivered by program staff according to the program design?	x	x	X	x
Content: Did teachers receive the intervention content as planned?	x	x	X	x
Modifications to original design: What were the primary adaptations and modifications made to the intervention as designed?	x	x	X	x
What factors that may have affected program fidelity and teacher participation?	x	x	x	x

Source: Authors.

Qualitative findings were based on themes that reoccurred across all implementation data sources. Qualitative data from these sources are not intended to be quantifiable, to be generalizable to the entire treatment group, to represent the views of all program staff, or to determine causal inferences or to draw conclusions about the impact or exploratory analysis presented in this report. Instead, these data are intended to provide insight into the expectations and experiences of the teachers and program staff who participated in Pacific CHILD and the factors that may have affected program fidelity and teacher participation.

Appendix L: Instruments

This appendix includes the following instruments from the study:

L.1 Teacher knowledge assessment

L.2 Additional eight items for the modified SIOP

L.3 Teacher background survey

L.4 Principal survey: Treatment group schools

L.5 Principal survey: Control group schools

L.6 Pacific CHILD interview guide: Program developer

L.7 Pacific CHILD interview guide: Project advisors

L.8 Pacific CHILD interview guide: Trainers

L.9 Teacher focus group discussion guide

L.10 Professional development observation summary for Annual Institute Week 1

L.11 Professional development observation summary for Annual Institute Week 2

L.12 Professional development observation summary for Mini Institute full day

L.13 Professional development observation summary for Mini Institute school-based day

L.14 Program staff lesson demonstration observation summary

L.15 Teacher lesson demonstration observation summary

L.16 Structured Learning Team observation (SLT)

L.1 Teacher knowledge assessment

Your school: _____

For questions 1 to 40, please circle [select] what you think is the best response for each question. Choose only one response for each question.

I. Instructional techniques and classroom environment

1. _____ is a complex network of existing knowledge.
    a. Scaffolding
    b. Cognition
    c. Proficiency
    d. Schema
    e. Revision

2. _____ includes the gradual withdrawal of teacher support.
    a. Comprehension
    b. Metacognition
    c. Assessment
    d. Scaffolding
    e. Engagement

3. Flexible grouping of students should be based on _____.
    a. Ethnicity/shared language
    b. Academic ability level
    c. Language proficiency level
    d. Standardized assessment data
    e. Classroom assessment data

4. _____ is when readers think about their comprehension processes as they read.
    a. Comprehension monitoring
    b. Word consciousness
    c. Reading fluency
    d. Engaged reading
    e. Instructional grouping

5. A teacher asks students to work together and assigns the following roles: facilitator, recorder, timekeeper, and reporter. This instructional technique is best characterized as:
    a. Role playing
    b. Interactive tasks
    c. Direct instruction
    d. Pull-out instruction
    e. Differentiated instruction

6. _____ is when readers think about their thinking.
    a. Metalinguistics
    b. Cognition
    c. Fluency
    d. Metacognition
    e. Acquisition

7. The **first step** toward building students' reading vocabulary involves building _____ vocabulary.
    a. Receptive
    b. Academic/technical
    c. Sight word
    d. Productive
    e. Oral

8. A teacher writes a list of words on the board: (1) ***as a result of***; (2) ***on the other hand***; and (3) ***the problem is***. These words will help students identify and break down the text structure of their reading. They are examples of:
    a. Graphic organizers
    b. Foundational words
    c. Content words
    d. Question generation
    e. Signal words

9. _____ always involve(s) carefully structured small-group activities that involve individual accountability, along with incentives for working well as a group and helping each other.
    a. Cooperative learning
    b. Question generation
    c. Story grammar
    d. Interactive tasks
    e. Differentiated instruction

10. Which of the following is an example of an effective use of "modeling"?
    a. Assessing prior knowledge
    b. Showing video clips to emphasize a particular concept
    c. Having students create a poster that summarizes a reading
    d. Having students orally repeat a set of vocabulary words to improve pronunciation
    e. Demonstrating a science experiment

11. _____ consist(s) of thinking procedures that guide readers when they are reading and writing so they know if they understand.
    a. Message redundancy
    b. Comprehensible input
    c. Comprehension strategies
    d. Text structure
    e. Interactive tasks

12. All of the following are important functions of a cognitively rich environment in the classroom EXCEPT:
    a. Supporting vocabulary acquisition
    b. Promoting independence
    c. Building background knowledge
    d. Addressing key standards and benchmarks
    e. Encouraging active student participation

13. In a lesson on cause and effect, a teacher is most likely to highlight the following phrases:
    a. ***in order to*** and ***therefore***
    b. ***defined as*** and ***for instance***
    c. ***on the other hand*** and ***however***
    d. ***different from*** and ***similar to***
    e. ***although*** and ***whereas***

14. _____ helps students build understanding of the text they are reading and connect those ideas to their past experiences and knowledge.
    a. Discrete skills mastery
    b. Summative assessment
    c. Explicit instruction
    d. Activating prior knowledge
    e. Formative assessment

15. _____ requires students to determine what is important about what they are reading and to briefly explain this information in their own words.
    a. Cognition
    b. Summarization
    c. Scaffolding
    d. Revision
    e. Read-aloud

16. In the ESL/ELD classroom, the **primary purpose** of interactive tasks is to _____.
    a. Eliminate the need for individual tasks
    b. Pre-teach the language required for content work
    c. Involve ELLs in meaningful content work
    d. Create homogeneous groupings
    e. Differentiate instruction for ELLs

17. Differentiated Instruction refers to _____
    a. Students learning by engaging in productive activity
    b. Teachers beginning a lesson at a predetermined set point
    c. Teachers sharing class materials to teach the same lesson
    d. Students learning from their textbooks, homework and tests
    e. Teachers adjusting instruction to meet individual needs

18. A teacher assigns a short text on whales and humans and highlights the phrases ***as a result*** and ***consequently***. The teacher is most likely focusing on:
    a. Oceanography
    b. Cause and effect
    c. Content vocabulary
    d. Compare and contrast
    e. Marine biology

19. Which of the following is NOT an essential component of vocabulary instruction?
    a. Providing rich and varied language experiences
    b. Explicitly teaching individual words
    c. Assessing spelling weekly
    d. Fostering word consciousness
    e. Teaching word-learning strategies

20. Instructional technique most clearly aligned with a socio-cultural view of learning is _____.
    a. Cross-age tutoring
    b. Direct instruction
    c. Differentiated instruction
    d. Scaffolding
    e. Sheltered instruction

21. Which of the following questions would best help to teach prediction during reading?
    a. Who is the author of the story?
    b. Who are the main characters?
    c. What will happen next?
    d. What did you like about the story?
    e. What is the author's purpose?

22. Why would a teacher activate students' prior knowledge before reading? In order to:
    a. Assess students' reading level.
    b. Increase students' reading fluency.
    c. Promote word consciousness.
    d. Develop students' sense of story structure.
    e. Make connections to the text

23. Creating charts of students' questions and modeling self questioning through thinking aloud are some of the ways teachers can encourage:
    a. Interactive tasks
    b. Question generation
    c. Message redundancy
    d. Story summarization
    e. Differentiated instruction

24. To help her students think about how the story they're reading is structured, the teacher says to the class, "Hmm, I am noticing a pattern here in this book." This is an example of:
    a. Independent reading
    b. Differentiated instruction
    c. Semantic mapping
    d. Modeling "think aloud"
    e. Question generation

25. Students are learning about volcanoes and earthquakes. Which foundational word might appear in their reading?
    a. Disaster
    b. Caldera
    c. Lithosphere
    d. Magma
    e. Seismic

26. Which of the following is a key characteristic of differentiated instruction?
    a. Assessment comes at the end and is summative
    b. Assignment options are the same for everyone
    c. Established curriculum and standards guide instruction
    d. Student interests and learning styles are the basis of instruction
    e. Differences are seen as problems

27. _____ involve(s) teaching students to flexibly use several different strategies and apply different strategies at different points in the text.
    a. Content focused instruction
    b. Multiple strategy instruction
    c. Flexible grouping strategies
    d. Cooperative group work
    e. Incidental word learning

II. Theories of Language Acquisition

28. Academic language refers to _____.
    a. Language that is used by a teacher or instructor
    b. Language used in social settings
    c. Written language like that used in college texts
    d. Language used in formal contexts for academic subjects
    e. highly complex language structures

29. _____ is the active construction of meaning from text.
    a. Reading fluency
    b. Reading comprehension
    c. Vocabulary acquisition
    d. Metacognitive awareness
    e. Skills mastery

30. The term morphology refers to _____.
    a. Explicit instruction in individual words
    b. The study of language change
    c. The study of context clues
    d. The study of word formation
    e. Explicit instruction in phonics

31. Metalinguistic knowledge involves the ability to_____.
    a. speak multiple languages
    b. Find hidden meanings in the text
    c. Talk about language forms and functions
    d. Connect new texts with prior knowledge
    e. Translate texts accurately

32. The best way to organize instruction for **intermediate to advanced** English Language Learners is to_____.
    a. Use simple sentences and below grade-level texts
    b. Ensure that students reach English proficiency before teaching grade level content
    c. Provide a specialized all-day program until ELLs reach oral fluency in English
    d. Begin with simple vocabulary and gradually add more complex vocabulary
    e. Use grade level curricula with appropriate support and scaffolding

33. The Zone of Proximal Development is the _____.
    a. Frustration level of the student
    b. Level at which the student is no longer progressing
    c. Difference between the level at which a learner can complete a task independently and the level at which she can complete it with support
    d. Difference between a learner's frustration level and acceptance level
    e. Level at which a student is able to work independently

f.

34. _____ should focus on helping students to engage in active processing.
    a. Question generation
    b. Guided practice
    c. Explicit instruction
    d. Academic language
    e. Multiple exposures

35. Cummins' Cognitive Academic Language Proficiency (CALP) is _____.
    a. The ability to engage in problem-solving, deduction, and complex memory tasks
    b. The level at which students are ready to be mainstreamed
    c. The language required to succeed in higher order, literacy-related tasks of the classroom
    d. The ability to use language in all its forms as a tool for thinking and communicating effectively
    e. When the primary language is partially or completely lost as a second language is acquired

36. Which formative assessment practice is best for understanding the current needs of a student learning to read?
    a. Administer a standardized reading test to students
    b. Have students engage in sustained silent reading
    c. Review students' performance on last year's reading test
    d. Observe the student reading a story out loud
    e. Assess the student's writing

37. Instructional conversation is an effective means for engaging ELLs in classroom discourse because it _____.
    a. Supports students so they make fewer mistakes
    b. Enables language learners to memorize correct forms
    c. Allows for students and teacher to follow a prepared script
    d. Provides different opportunities for modeling and feedback that support language learning
    e. Prevents students from repeating each other's errors

38. According to Krashen, comprehensible input is _____.
    a. The order in which certain features of a language are acquired
    b. Language input just beyond students' current ability level
    c. Language input simplified to students' current ability level
    d. Translation into the first language to ensure comprehension
    e. The interaction of emotional factors with other factors that affect comprehension

39. Leaving out elements of a sentence is called a(n)_____ error.
    a. Systematic
    b. Developmental
    c. Overgeneralization
    d. Transfer
    e. Simplification

40. Additive Bilingualism is _____.
    a. Developing a student's primary language while he or she acquires a second language
    b. The ability to engage in problem-solving, deduction, and complex memory tasks
    c. Having equal proficiency in two languages
    d. The act of acquiring a third or fourth language
    e. Replacing the primary language with the second language

L.2 Additional eight items for the modified SIOP

31. SIOP-PLUS ITEM: Text Structure				
4	**3**	**2**	**1**	**0**
Teacher provides explicit instruction on **text structure** (clearly defines text structure and provides an explanation on how text structure assists with understanding the text.)		Instruction **on text structure** is implied.		Teacher does not provide any instruction **on text structure** or the curriculum did not provide the opportunity for the teaching of **text structure**.
Notes: Text structure refers to the organizational pattern an author uses to structure ideas in a text. Common text structures include: Sequence, Compare/Contrast, Cause/Effect, Problem/Solution *Examples of explicit text structure: Teacher defines structure, provides examples, reviews signal words, asks questions about the text structure, and connects visual representations to patterns in text of specific text structures.*				
Comments:				

32. SIOP-PLUS ITEM: Vocabulary/Word Learning Strategies				
4	**3**	**2**	**1**	**0**
Teacher provides clearly defined **word learning strategies** (e.g. an explicit description of the strategy and how it should be used, context clues, word parts, or the explicit use of resources to deepen student's understanding of word meanings).		**Word learning strategies** are implied. There is no clearly defined link between instruction in vocabulary and word learning strategies.		Teacher does not provide any instruction on **word learning strategies**.
Notes: *Word learning strategies include the following:* ▪ ***Context Clues*** *include words, phrases, and sentences that provide clues to the meaning of unknown words.* ▪ ***Word Parts*** *include prefixes, roots, suffixes, and affixes.* ▪ ***Resources****: the use of a dictionary, thesaurus, glossary, or encyclopedia to learn word meanings and to deepen knowledge of word meanings.*				
Comments:				

33. SIOP-PLUS: Student Question Generation (Quality)				
4	**3**	**2**	**1**	**0**
Students ask **a range of questions** during the language arts lesson (literal, inferential, evaluative)		Students ask **only literal questions** during the language arts lesson.		Students **do not ask any questions** during the language arts lesson.

Notes:
Includes questions students ask teachers, questions students write down, and questions that students ask of one another during group work.

- ***Literal:*** *The answer is directly and clearly stated in the text. Example: Where do sea turtles live?*
- ***Inferential****: The answer requires the reader to combine background knowledge with details in the text to form a conclusion or interpretation. Example: Why are sea turtles endangered?*
- ***Evaluative****: The answer is not found in the text. The reader has to use his/her own background and make an opinion or a judgment about the information in the text.*

Comments:

34. Student Question Generation (Quantity)				
4	**3**	**2**	**1**	**0**
Most students ask questions during the language arts lessons		**Few students** ask questions during the language arts lesson.		**No students** ask questions during the language arts lesson

Comments:

35. SIOP-PLUS: Cognitively Rich Environment				
4	**3**	**2**	**1**	**0**
There is **ample evidence** of instruction in reading skills in the classroom		There is **some evidence** of instruction in reading skills in the classroom		There is **no evidence** of instruction in reading skills in the classroom

Notes:
Examples: Student and teacher work is posted, print on walls, classroom library, signal words, realia, references such as encyclopedias, dictionaries, and thesauruses, newspapers, magazines, word walls, sentence strips
visual representations, including graphic organizers, charts, diagrams, graphs, and maps.

Comments:

36. SIOP-PLUS: Differentiated Instruction				
4	**3**	**2**	**1**	**0**
Teacher **effectively differentiates** content, process, and/or product based on students' needs		Teacher **attempts to differentiate** content, process, and/or product based on students' needs		Teacher **does not attempt to differentiate** content, process, and/or product based on students' needs

Notes:
Differentiated instruction is an educational approach that adjusts instruction to accommodate individual students' needs, rather than beginning at a predetermined set point.

- ***Content/Topic****: What is to be learned*
- ***Process****: How to facilitate learning*
- ***Product:*** *What students did and learned*
- ***Student needs****: Readiness (ELP and background), learning styles, and interests.*

Comments:

37. SIOP-PLUS: Flexible Groups				
4	**3**	**2**	**1**	**0**
Teacher places students in flexible groups that support their learning		Teacher places students in flexible groups, but it is not clear how these groups support or scaffold learning (it is not clear how the grouping is connected to the activity)	Teacher places students in groups, but they are not flexible	Teacher does not place students in groups

Notes:

- *Flexible groups are temporary groups assigned for a specific task. Though groups may sometimes be based on skill level, they are designed to provide cooperative peer support to students in the completion of a specific step in a larger assignment. These are not ability groups designed for the teacher to provide levelized small group instruction.*
- *Flexible groups provide opportunities for students to be members of* ***more than one group****. Students can be grouped and regrouped according to specific goals, activities, and individual needs.*

Comments:

38. SIOP-PLUS: Individual Accountability				
4	**3**	**2**	**1**	**0**
The majority of the students demonstrate individual accountability in the group as students work toward a common goal		A few students demonstrate individual accountability in the group as students work toward a common goal		Students do not demonstrate individual accountability in the group as students work toward a common goal. OR Students are not placed in groups at any time during the lesson
Comments:				

L.3 Teacher background survey

I. Your Professional and Demographic Background

1. Where do you teach?

School: _____

Grade level:_____

2. Which of the following do you currently hold? (Check all appropriate boxes)

- ☐ High school diploma
- ☐ Associate's degree
- ☐ Bachelor's degree
- ☐ Master's degree
- ☐ Doctorate or professional degree (Ph.D., Ed.D., M.D., J.D., D.D.S., etc.)
- ☐ Other (Specify) _____
- ☐ None of the above

3. Which of the following teaching credentials or professional licenses do you hold? (Check all appropriate boxes)

- ☐ Teaching Credential
  - ❒ Hawai'i State
  - ❒ American Samoa
  - ❒ CNMI
  - ❒ Other state/entity (Specify state or entity): _____
- ☐ Special Endorsement
  - ❑ Special Education
  - ❑ ESL/TESOL
  - ❑ Bilingual Education
  - ❑ Other (Specify type) _____
- ☐ Administrator
- ☐ Substitute/Provisional License
- ☐ National Board Certification
- ☐ Other (please specify) _____
- ☐ None of the above

4. What is your gender?

- ☐ Male
- ☐ Female

5. Which of the following best describes your racial or ethnic background? (Check all appropriate boxes)

- ☐ American Indian or Alaska Native
  - ❑ Specify Tribe or Nation _____
- ☐ Black/African American
- ☐ Asian

  Please specify
  - ❑ Chinese
  - ❑ Indian
  - ❑ Japanese
  - ❑ Korean
  - ❑ Other (Specify) _____
- ☐ Pacific Islander or Native Hawaiian
  - ❑ Carolinian
  - ❑ Chamorro
  - ❑ Chuukese
  - ❑ Filipino
  - ❑ Kapingese
  - ❑ Kosrean
  - ❑ Marshallese
  - ❑ Mwokliese
  - ❑ Native Hawai'ian
  - ❑ Ngatikese
  - ❑ Nukuoran
  - ❑ Palauan
  - ❑ Pingelapese
  - ❑ Pohnpeian
  - ❑ Samoan
  - ❑ Satawalese
  - ❑ Tongan
  - ❑ Ulithian
  - ❑ Yapese
  - ❑ Other (Specify) _____
- ☐ Hispanic/Latino
- ☐ White
- ☐ Other (Specify)_____

6. **What is your dominant language (the language you feel you speak best or are the most comfortable using)?**

   ☐ English (standard)
   ☐ Pidgin
   ☐ Samoan
   ☐ Carolinian
   ☐ Chamorro
   ☐ Other: Specify _____

7. **Do you have conversational fluency in <u>any</u> language(s) or dialect(s) other than your dominant language or English? (This includes Pidgin.)**

   ☐ No (Skip to Question #9)
   ☐ Yes

8. **If you selected "Yes" in Question 7 above, please specify the language(s) (including Pidgin):**

   _____

9. **How do you rate your English fluency?**

   ☐ Native fluency, and English is my dominant language
   ☐ Near native fluency, but English is not my dominant language
   ☐ Advanced fluency (can read, write and conduct business in English with no or little difficulty)
   ☐ Intermediate fluency (can read, write and conduct business in English with occasional difficulty)
   ☐ Less than intermediate fluency

10. **Please tell us about your past teaching experience. Not counting the current school year, how many years have you been teaching at *any K-12* school (either full time or part time)? Include all teaching experience except student teaching.**

_____ year(s) (If less than one year, _____ months)

11. **Not counting the current school year, how many years have you been teaching at your current school? Include all teaching experience except student teaching.**

_____ year(s) (If less than one year, _____ months)

12. **Of all your years spent teaching, how many years did you spend teaching in classes in which 20 percent or more students (i.e., at least one in five students) were English Language Learners (as defined by your district/DOE/PSS)?**

_____ year(s)

13. **What is the approximate total number of students in your 4th or 5th grade class this year?**

_____ students

14. **Of the total number of students listed above, approximately how many are English Language Learners (as defined by your district/DOE/PSS)?**

_____ ELLs

15. **[Program Group School Version] Other than Pacific CHILD, please indicate to what extent you have participated in professional development focused on the following topics, during summer {20XX} and school year {20XX}-{20YY}.**

[Control Group School Version] Please indicate to what extent you have participated in professional development focused on the following topics, {20XX} and school year {20XX}-{20YY}.

(Do not include Pacific CHILD activities.)	Not at All	1-3 Days	4-6 Days	7 Days or More
a. English language development standards	☐	☐	☐	☐
b. Content area standards	☐	☐	☐	☐
c. Support for a published curriculum in language arts or other content areas	☐	☐	☐	☐
d. Differentiated instruction	☐	☐	☐	☐
e. PRExC-ELL or ExC-ELL® Training	☐	☐	☐	☐
f. Direct Instruction (DI) training	☐	☐	☐	☐
g. Success for All training	☐	☐	☐	☐
h. Corrective Reading (CR) training	☐	☐	☐	☐
i. Open Court Training	☐	☐	☐	☐
j. The Sheltered Instruction Observation Protocol (SIOP) training	☐	☐	☐	☐
k. Other training on instructional techniques Please specify: _____	☐	☐	☐	☐
l. Other training not mentioned above: Please specify: _____	☐	☐	☐	☐

II. Instructional Practices

16. Please indicate how often you use the following instructional techniques in your classroom. (Check one box per row)

	Never/ Almost Never	1-2 Times per Month	1-2 Times per Week	1-2 Times per Day	Several Times per Day
a. Explicitly teach academic language particular to English Language Arts	☐	☐	☐	☐	☐
b. Use multiple techniques to make concepts and tasks clear (e.g., visuals, manipulatives, realia, modeling)	☐	☐	☐	☐	☐
c. Provide below-grade-level materials available for students with lower English proficiency	☐	☐	☐	☐	☐
d. Provide opportunities for all students to use higher-order thinking skills (e.g., problem solving, predicting, organizing, evaluating, self-monitoring)	☐	☐	☐	☐	☐
e. Simplify language input to make it more comprehensible to English language learners	☐	☐	☐	☐	☐
f. Use the students' primary language to clarify concepts	☐	☐	☐	☐	☐
g. Adjust expectations for students whose limited English proficiency prevents them from meeting instructional targets	☐	☐	☐	☐	☐
h. Provide students with extra wait time	☐	☐	☐	☐	☐
i. Explicitly teach reading comprehension strategies	☐	☐	☐	☐	☐
j. Group students by their proficiency in English	☐	☐	☐	☐	☐
k. Create groups of students, each consisting of students with the <u>same</u> skill/comprehension levels	☐	☐	☐	☐	☐
l. Create groups of students, each consisting of students with <u>different</u> skills/comprehension levels.	☐	☐	☐	☐	☐
m. Explicitly correct student speaking errors (e.g., pronunciation, grammar)	☐	☐	☐	☐	☐

17. How frequently do you typically ask your students, <u>including</u> English language learners, to engage in the following activities? (Check one box per row)

	Never/ Almost Never	1-2 Times per Month	1-2 Times per Week	1-2 Times per Day
a. Develop oral or written summaries of reading	☐	☐	☐	☐
b. Evaluate their own work	☐	☐	☐	☐
c. Complete workbook or textbook exercises in class	☐	☐	☐	☐
d. Evaluate a piece of work completed by another student	☐	☐	☐	☐
e. Memorize vocabulary, facts, rules, or procedures	☐	☐	☐	☐
f. Engage in discussions about a reading	☐	☐	☐	☐
g. Listen to lectures and take notes	☐	☐	☐	☐
h. Work in small groups of two or more students	☐	☐	☐	☐
i. Recite poetry, speeches, or passages from memory	☐	☐	☐	☐
j. Use data and text references to support their ideas	☐	☐	☐	☐
k. Complete tests or quizzes	☐	☐	☐	☐

18. To what extent is each of the following a challenge at your school?
(Check one box per row)

	Not a Challenge	Minor Challenge	Moderate Challenge	Serious Challenge
a. Shortage of certified teachers	☐	☐	☐	☐
b. Shortage of ESL or bilingual teachers	☐	☐	☐	☐
c. Time for teachers to collaborate	☐	☐	☐	☐
d. A high proportion of English language learners	☐	☐	☐	☐
e. Student behavior/discipline	☐	☐	☐	☐
f. Lack of community or parent support	☐	☐	☐	☐
g. Lack of student motivation	☐	☐	☐	☐
h. Lack of appropriate materials for ELLs	☐	☐	☐	☐
i. Collegiality among faculty	☐	☐	☐	☐
j. Lack of administrative support	☐	☐	☐	☐
k. Lack of professional development opportunities	☐	☐	☐	☐

19. Please indicate the extent to which you agree or disagree with the following statements about student learning. (Check one box per row)

	Strongly Disagree	Disagree	*Agree*	Strongly Agree
a. Some students I teach are not able to learn the material I am supposed to teach them	☐	☐	☐	☐
b. ELL students should develop and maintain their primary language	☐	☐	☐	☐
c. Teachers should incorporate the cultures of their students into instructional activities	☐	☐	☐	☐
d. The use of primary language in the classroom slows down English language learning	☐	☐	☐	☐
e. Teachers should modify the curriculum to meet the needs of ELL students	☐	☐	☐	☐
f. ELL students succeed in mainstream classes with native English speaking peers	☐	☐	☐	☐
g. The use of primary language at home can impede learning a second language	☐	☐	☐	☐
h. Students may use their primary languages in my classroom	☐	☐	☐	☐
i. Culture is part of the curriculum in my classroom	☐	☐	☐	☐
j. I feel I have the preparation or professional development necessary to meet the needs of ELL students	☐	☐	☐	☐
k. The presence of ELL students in mainstream classrooms has a negative impact on the achievement of other students	☐	☐	☐	☐
l. The best way to deal with an ELL's lack of comprehension is to use simplified materials and simplified language	☐	☐	☐	☐

20. To what extent do you agree with the following statements? (Check one box per row)

" My ability to teach to the Language Arts standards is limited by . . . . ."

	Not at All	Very Little	Somewhat	*A Great Deal*
a. The number of English language learners in my class	☐	☐	☐	☐
b. The low ability of my students	☐	☐	☐	☐
c. The level of parent or community support	☐	☐	☐	☐
d. My knowledge of working with English language learners	☐	☐	☐	☐
e. The range of students' needs in my class	☐	☐	☐	☐
f. A lack of support from principals/administrators	☐	☐	☐	☐
g. My knowledge of my content area	☐	☐	☐	☐
h. A lack of support from other teachers	☐	☐	☐	☐
i. Availability of materials and resources	☐	☐	☐	☐
j. My lack of fluency in the English language	☐	☐	☐	☐
k. The requirement that instruction be ONLY in English	☐	☐	☐	☐

III. Pacific CHILD Professional Development Support

[Note: Section III Questions 21-31 apply only to the program group school teachers.]

21. Did you participate in the 10-day Summer Institute in the summer of {20XX}?

☐ Attended all days of the training [Skip to Question23]

☐ Attended 6-9 days of the training

☐ Attended 1-5 days of the training

☐ I was invited to participate, but did not attend

☐ N/A (I joined the Pacific CHILD Program after the Summer Institute). **[Skip to Question 23]**

22. If you attended fewer than 10 days of the {20XX} Summer Institute, what are the reasons that you did not participate or missed days of the training? (Check all that apply)

☐ Schedule conflict with my personal plans

☐ Schedule conflict with my school/professional activities

☐ School/district did not compensate me for my attendance in the training

☐ Did not feel the Pacific CHILD model was culturally appropriate for my students

☐ Did not think that the summer institute would be useful for me

☐ Other: Please explain: _____

23. How many follow-up mini institutes have you attended so far during the {20XX}-{20YY} school year? Did you attend all mini institutes offered so far?

Number of mini-institutes attended this year (Circle one): 0 1 2 3

☐ Attended all mini institutes offered so far (did not miss any mini-institute that was offered) **[Skip to Question 25]**

☐ Did not attend any or some of mini institutes offered so far.

☐ N/A (No mini-institutes have been offered to me so far.) **[Skip to Question 25]**

24. What are the main reasons that you did not participate in or missed a mini-institute? (Check all that apply)

☐ Schedule conflict with my personal plans

☐ Schedule conflict with my school/professional activities

☐ School/district did not compensate me for my attendance in the training

☐ Did not think that the mini-institute would be useful for me

☐ Other: Please explain: _____

25. How often did you attend the weekly Structured Learning Team meeting?

☐ 3-4 times a month **[Skip to Question 27]**

☐ 2-3 times a month

☐ 1-2 times a month

☐ 1-2 times per semester

☐ 1 time or not at all

26. What are the main reasons that you did not regularly attend the weekly Structured Learning Team meeting?

☐ Schedule conflict with my personal plans
☐ Schedule conflict with my school/professional activities
☐ School/district did not compensate me for my attendance in the meeting
☐ Did not think that the Structured Learning Team would be useful for me
☐ Other: Please explain: _____

27. How would you describe your participation in the twice monthly classroom observations and monthly demonstration lessons with the PREL Pacific CHILD training staff?

☐ Met with PREL staff about twice a month **[Skip to Question 29]**

☐ Met with PREL staff about once a month

☐ Met with PREL staff less frequently than once a month

☐ Did not meet with PREL staff at all

28. What are the main reasons that you did not meet regularly with PREL staff for the twice monthly classroom observations and monthly demonstration lessons?

- ☐ Schedule conflict with my personal plans
- ☐ Schedule conflict with my school/professional activities
- ☐ School/district did not compensate me for my attendance in the meeting
- ☐ Classroom observations and demonstrations were not regularly scheduled at my school
- ☐ Did not think that the meeting with PREL staff would be useful for me
- ☐ Other: Please explain: _____

29. How would you best describe the availability of your PREL training staff?

- ☐ My PREL staff was frequently present at my school and tried to meet with me often
- ☐ My PREL staff was sometimes present at my school and tried to meet with me sometimes
- ☐ My PREL staff was rarely present at my school and tried to meet with me occasionally
- ☐ My PREL staff was never present at my school and did not try to meet with me at all
- ☐ Other: Please explain: _____

30. Please indicate the extent to which you agree or disagree with each of the following statements:

	Strongly Disagree	Disagree	Agree	Strongly Agree	Don't Know
a. PREL staff encouraged me to collaborate with other teachers and work on instructional plans together	☐	☐	☐	☐	☐
b. PREL staff emphasized the Pacific CHILD components that were learned during the institutes	☐	☐	☐	☐	☐
c. PREL staff helped me improve my instructional plans	☐	☐	☐	☐	☐
d. After each Structured Learning Team meeting, I was asked to reflect on and evaluate the session	☐	☐	☐	☐	☐
e. PREL staff helped me to understand the purpose behind instructional practices	☐	☐	☐	☐	☐
f. PREL staff was available to me outside of the Structured Learning Team meeting time	☐	☐	☐	☐	☐
g. Structured Learning Team meetings were well thought out and organized	☐	☐	☐	☐	☐
h. PREL staff was knowledgeable with respect to the Pacific CHILD components	☐	☐	☐	☐	☐
i. PREL staff was knowledgeable with respect to implementing a variety of instructional practices	☐	☐	☐	☐	☐
j. PREL staff was knowledgeable with respect to implementing differentiated instruction	☐	☐	☐	☐	☐
k. I feel I will be able to continue using Pacific CHILD components after the study has ended	☐	☐	☐	☐	☐
l. The support of my PREL staff was valuable to my professional development	☐	☐	☐	☐	☐

31. Please indicate the extent to which you feel the following Pacific CHILD activities were useful:

	Not Useful At All	Not Very Useful	Somewhat Useful	*Very Useful*	*These activities were not provided or discussed*
a. Gathering information about my teaching context	☐	☐	☐	☐	☐
b. Identifying learning goals for my instructional plans	☐	☐	☐	☐	☐
c. Identifying activities that will support the learning goals	☐	☐	☐	☐	☐
d. Discussing grade–appropriate academic content	☐	☐	☐	☐	☐
e. Discussing language learning issues	☐	☐	☐	☐	☐
f. Pre-conference conversation before the observation	☐	☐	☐	☐	☐
g. Discussing how my instructional plan supports all students' achievement	☐	☐	☐	☐	☐
h. Having my instructional practice observed by the PREL staff	☐	☐	☐	☐	☐
i. Receiving feedback on my teaching from my PREL staff after the observation	☐	☐	☐	☐	☐
j. PREL staff's demonstrations of instructional practices	☐	☐	☐	☐	☐
k. The questions my PREL staff asked to help me reflect on specific aspects of my teaching	☐	☐	☐	☐	☐
l. Revising/identifying the instructional goals for my future plans	☐	☐	☐	☐	☐

32. In your experience how effective is Pacific CHILD in the following areas:

		Not Effective at All	Not Very Effective	Somewhat Effective	Very Effective	Don't Know
a.	Your knowledge regarding the needs of English language learners	☐	☐	☐	☐	☐
b.	Your confidence in implementing the Pacific CHILD components in the classroom	☐	☐	☐	☐	☐
c.	Your awareness of the cultures of English language learners	☐	☐	☐	☐	☐
d.	Your knowledge regarding reading comprehension strategies	☐	☐	☐	☐	☐
e.	Your ability to differentiate instruction	☐	☐	☐	☐	☐
f.	Your use of vocabulary development techniques with students	☐	☐	☐	☐	☐
g.	Your use of flexible grouping strategies	☐	☐	☐	☐	☐
h.	Your ability to create a cognitively rich environment in the classroom	☐	☐	☐	☐	☐
i.	Your knowledge of English Language Development	☐	☐	☐	☐	☐
j.	Your ability to incorporate interactive tasks into your lessons.	☐	☐	☐	☐	☐

L.4 Principal survey: Treatment group schools

Today's Date: _____/_____/_____
Day Month Year

School: _____

I. Number of Teachers in Your School

1. Please provide the following information about your current 4th grade teachers.

a. Number of 4th grade teachers (current)	
b. Number of 4th grade teachers who are new or transferred to your school this year (2009-10) to date	
c. Number of 4th grade teachers you expect will return to teach the 4th grade next year	
d. Number of 4th grade teachers you expect will return but will teach a different grade next year	
e. Number of 4th grade teachers you expect to leave or retire from the school at the end of this school year (2009-10)	

2. Please provide the following information about your current 5th grade teachers.

a. Number of 5th grade teachers (current)	
b. Number of 5th grade teachers who are new or transferred to your school this year (2009-10) to date	
c. Number of 5th grade teachers you expect will return to teach the 5th grade next year	
d. Number of 5th grade teachers you expect will return but will teach a different grade next year	
e. Number of 5th grade teachers you expect to leave or retire from the school at the end of this school year (2009-10)	

3. How many reading/ELL/ESL specialists, if any, currently work with 4th and/or 5th graders?

_____ Specialists

II. Number of Students in Your School

4. Please provide the following information about your current 4th grade students.

a. Number of 4th grade students	
b. Number of 4th grade ELL/ESL students	
c. Number of 4th grade Special Education students	
d. Number of 4th grade students transferred to your school during this school year (2009-2010) to date	
e. Number of 4th grade students who left your school during this school year (2009-2010) to date	

5. Please provide the following information about your current 5th grade students.

a. Number of 5th grade students	
b. Number of 5th grade ELL/ESL students	
c. Number of 5th grade Special Education students	
d. Number of 5th grade students transferred to your school during this school year (2009-2010) to date	
e. Number of 5th grade students who left your school during this school year (2009-2010) to date	

III. Policies Regarding Professional Development of Teachers

6. How many hours or days of professional development are the 4th and 5th grade classroom teachers required to complete during the school year 2009-2010. Please choose one:

☐ At least _____ Hours OR _____ Days

☐ There is no required or expected minimum hours

7. What type of professional development activities are the 4th and 5th grade teachers at your school required by the state/district to participate in during the school year 2009-2010? (Do not include Pacific CHILD activities)

Subject Area	Required for all teachers	Required for some teachers	Not required but recommended for teachers	Not required and not a priority for teachers
a. English language development standards	☐	☐	☐	☐
b. Content area standards	☐	☐	☐	☐
c. Support for a published curriculum in language arts or other content areas	☐	☐	☐	☐
d. Differentiation of instruction	☐	☐	☐	☐
e. Direct Instruction training	☐	☐	☐	☐
f. Corrective Reading training	☐	☐	☐	☐
g. The Sheltered Instruction Observation Protocol (SIOP) training	☐	☐	☐	☐
h. Open Court Training	☐	☐	☐	☐
i. Success For All Training	☐	☐	☐	☐
j. Other training on instructional techniques. Please specify: _____	☐	☐	☐	☐
k. Other training not mentioned above. Please specify: _____	☐	☐	☐	☐

IV. Challenges at School

8. What do you see as the main challenges facing your 4th and 5th grade teachers? Please check the box that best describes your observation.

	Serious Challenge	Moderate Challenge	Minor Challenge	Not a Challenge
a. Lack of ESL certification	☐	☐	☐	☐
b. Lack of bilingual certification	☐	☐	☐	☐
c. Lack of time for teachers to collaborate	☐	☐	☐	☐
d. High proportion of English language learners	☐	☐	☐	☐
e. Student behavior/discipline	☐	☐	☐	☐
f. Lack of community or parent support	☐	☐	☐	☐
g. Lack of student motivation	☐	☐	☐	☐
h. Lack of appropriate materials for ELLs	☐	☐	☐	☐
i. Degree of collegiality among faculty	☐	☐	☐	☐
j. Lack of coordination/communication with administration	☐	☐	☐	☐
k. Lack of professional development opportunities	☐	☐	☐	☐

9. Some or all of your 4th and 5th grade teachers are participating in the Pacific CHILD Professional Development Program this year. How would you describe the participation of these teachers at your school in Pacific CHILD? Please select the response that best describes your observations.

☐ All or most of the teachers actively and consistently participate in Pacific CHILD (e.g., attend training institutes, participate in weekly peer-support activities, etc.)

☐ Some of the teachers actively and consistently participate in Pacific CHILD

☐ Few of the teachers participate actively or consistently in Pacific CHILD

☐ I do not know / I am not sure

10. If a teacher is not participating actively or consistently in the Pacific CHILD, what do you think are the most important reasons for this? Please choose up to two reasons (one primary reason and one secondary reason).

	Primary Reason (check one box)	Secondary Reason (check one box)
a. Teacher declined to participate from the start as his/her personal choice	☐	☐
b. Teacher is too busy to make time for Pacific CHILD activities	☐	☐
c. Teacher feels that the program is not helpful or useful	☐	☐
d. Teacher finds the program to be poorly implemented	☐	☐
e. Other (Please explain): _____	☐	☐
f. I do not know / I am not sure	☐	☐

11. How would you describe your support for Pacific CHILD?

☐ I fully support Pacific CHILD and encourage all 4th and 5th grade teachers to participate

☐ I fully support Pacific CHILD, but I take a neutral stance on the teachers' own decisions about participating in the program

☐ I support Pacific CHILD, but with a little reservation

☐ I support Pacific CHILD, but with considerable reservation

☐ I do not support Pacific CHILD

☐ I do not know / I am not sure

12. Do you have any concerns about the following issues regarding Pacific CHILD?

	I am concerned . . .					
	Very much	Somewhat	Not very much	Not at all	N/A	Don't know
a. Teachers' willingness and commitment to participate	☐	☐	☐	☐	☐	☐
b. Helpfulness /usefulness of the program	☐	☐	☐	☐	☐	☐
c. Resources needed to provide substitutes so that the teachers can participate in Pacific CHILD	☐	☐	☐	☐	☐	☐
d. Resources needed to pay for teachers' time while attending training	☐	☐	☐	☐	☐	☐
e. Time commitment by teachers to participate in Pacific CHILD (e.g., training institutes, mini institutes, weekly meetings)	☐	☐	☐	☐	☐	☐
f. Time commitment by teachers to participate in the evaluation research activities (e.g., surveys)	☐	☐	☐	☐	☐	☐
g. Other (Please list): _____	☐	☐	☐	☐	☐	☐

13. If you have any comments about any aspect of the Pacific CHILD program or its evaluation activities, please provide them below.

Thank you for your time.

L.5 Principal survey: Control group schools

Today's Date: _____/_____/_____
Day Month Year

School: _____

I. Number of Teachers in Your School

1. Please provide the following information about your current 4th grade teachers.

a. Number of 4th grade teachers (current)	
b. Number of 4th grade teachers who are new or transferred to your school this year (2009-10) to date	
c. Number of 4th grade teachers you expect will return to teach the 4th grade next year	
d. Number of 4th grade teachers you expect will return but will teach a different grade next year	
e. Number of 4th grade teachers you expect to leave or retire from the school at the end of this school year (2009-10)	

2. Please provide the following information about your current 5th grade teachers.

a. Number of 5th grade teachers (current)	
b. Number of 5th grade teachers who are new or transferred to your school this year (2009-10) to date	
c. Number of 5th grade teachers you expect will return to teach the 5th grade next year	
d. Number of 5th grade teachers you expect will return but will teach a different grade next year	
e. Number of 5th grade teachers you expect to leave or retire from the school at the end of this school year (2009-10)	

3. How many reading/ELL/ESL specialists, if any, currently work with 4th and/or 5th graders?
_____ Specialists

II. Number of Students in Your School

4. Please provide the following information about your current 4th grade students.

a. Number of 4th grade students	
b. Number of 4th grade ELL/ESL students	
c. Number of 4th grade Special Education students	
d. Number of 4th grade students transferred to your school during this school year (2009-2010) to date	
e. Number of 4th grade students who left your school during this school year (2009-2010) to date	

5. Please provide the following information about your current 5th grade students.

a. Number of 5th grade students	
b. Number of 5th grade ELL/ESL students	
c. Number of 5th grade Special Education students	
d. Number of 5th grade students transferred to your school during this school year (2009-2010) to date	
e. Number of 5th grade students who left your school during this school year (2009-2010) to date	

III. Policies Regarding Professional Development of Teachers

6. How many hours or days of professional development are the 4th and 5th grade classroom teachers required to complete during the school year (2009-2010). Please choose one:

☐ At least _____ Hours OR _____ Days

☐ There is no required or expected minimum hours

7. What type of professional development activities are the 4th and 5th grade teachers at your school required by the state/district to participate in during the school year (2009-2010)?

Subject Area	Required for all teachers	Required for some teachers	Not required but recommended for teachers	Not required and not a priority for teachers
a. English language development standards	☐	☐	☐	☐
b. Content area standards	☐	☐	☐	☐
c. Support for a published curriculum in language arts or other content areas	☐	☐	☐	☐
d. Differentiation of instruction	☐	☐	☐	☐
e. Direct Instruction training	☐	☐	☐	☐
f. Corrective Reading training	☐	☐	☐	☐
g. The Sheltered Instruction Observation Protocol (SIOP) training	☐	☐	☐	☐
h. Open Court Training	☐	☐	☐	☐
i. Success For All Training	☐	☐	☐	☐
j. Other training on instructional techniques. Please specify: _____	☐	☐	☐	☐
k. Other training not mentioned above. Please specify: _____	☐	☐	☐	☐

IV. Challenges at School

8. What do you see as the main challenges facing your 4th and 5th grade teachers? Please check the box that best describes your observation.

	Serious Challenge	Moderate Challenge	Minor Challenge	Not a Challenge
a. Lack of ESL certification	☐	☐	☐	☐
b. Lack of bilingual certification	☐	☐	☐	☐
c. Lack of time for teachers to collaborate	☐	☐	☐	☐
d. High proportion of English language learners	☐	☐	☐	☐
e. Student behavior/discipline	☐	☐	☐	☐
f. Lack of community or parent support	☐	☐	☐	☐
g. Lack of student motivation	☐	☐	☐	☐
h. Lack of appropriate materials for ELLs	☐	☐	☐	☐
i. Degree of collegiality among faculty	☐	☐	☐	☐
j. Lack of coordination/communication with administration	☐	☐	☐	☐
k. Lack of professional development opportunities	☐	☐	☐	☐

9. How many of your 4th and 5th grade teachers have participated or will participate in training which is aimed at <u>improving their skills to teach English Language Learners</u> this year (2009-2010)?

☐ All or most of the 4th and 5th grade teachers

☐ Some of the 4th and 5th grade teachers

☐ None of the 4th and 5th grade teachers **[Skip to Question 11]**

10. **If you indicated that your teachers are participating in training aimed at improving their skills to teach English Language Learners, please briefly describe this training:**

11. **How many of your 4th and 5th grade teachers have participated or will participate in <u>training aimed at improving their general teaching skills</u> this year (2009-2010)?**

- ☐ All or most of the 4th and 5th grade teachers
- ☐ Some of the 4th and 5th grade teachers
- ☐ None of the 4th and 5th grade teachers **[Skip to Question 13]**

12. **If you indicated that your teachers are participating in training aimed at improving general teaching skills, please briefly describe this training:**

13. **If you have any comments about any aspect of the Pacific CHILD evaluation activities, please provide them below.**

Thank you for your time.

L.6 Pacific CHILD interview guide: Program developer

Date _____

Interviewer Name _____

Respondent Title _____

Interviewer Instructions

Where respondent is unfamiliar with the topic of a question, please skip or tailor to respondents specific role in Pacific CHILD, i.e., trainers of trainers vs. administrators of other types. Remember that different persons have different kinds of knowledge about Pacific CHILD, its purpose, implementation, and impact and therefore need to be encouraged to answer in a manner commensurate with their role.

Opening Script

Thank you for meeting with us. The purpose of our discussion is to learn more about your experiences with Pacific CHILD. Your point of view as a Pacific CHILD program developer and project advisor is extremely valuable to us. We are especially interested in learning more about how you feel the professional development is going so far, and your impressions of teachers' progress in implementing the core elements of Pacific CHILD to work more effectively with ELLs.

We want today to focus on the implementation of Pacific CHILD during the 2008-2009 school year in American Samoa and CNMI (Year 2) and Hawaii (Year 1)

All of your answers to our questions today are entirely voluntary. We will maintain strict confidentiality for any answers you provide and your name will not be associated with any of your answers. Instead, your answers will be combined with those of others in our analyses and reporting, without your name attached to them, to help create a portrait of Pacific CHILD implementation.

We will record our conversation today for the purposes of note-taking only. All recordings and notes will be kept in secured, locked facilities for the duration of this project and destroyed upon project completion.

Do you have any questions before we begin?

Background

1. Please briefly tell me about your background and experience prior to working with PREL and/or on Pacific CHILD. (*Probe: years of teaching experience if any, experience with providing support to ELL teachers, other experiences developing teacher PD programs)*

2. How long have you been with PREL? How long have you worked on the Pacific CHILD program?

3. Please describe your role in Pacific CHILD during the 2008-2009 implementation of Pacific CHILD. (Probe: program design, manual revision, school recruitment, summer institutes, mini-institutes, demonstrations/observations, SLTs, school liaisons, etc.).

4. [*For those who identify a role in program design, please follow up with this question*:] Please describe your role in developing the content of the Pacific CHILD itself. How did you determine program content? What were your goals for the program? What did you hope the program would achieve? How has the content changed or modified for the 2008-2009 year (including the teacher and professional development manuals)?

5. Which components or aspects of Pacific CHILD do you consider to be essential to the model, irrespective of context (i.e., which components do you feel should not be modified or adapted)

6. How would you describe the role of (a) Project Advisors and (b) trainers (of teachers) in the Pacific CHILD program?

Trainers' Need for Training on Providing Professional Development

1. Please describe your role in working with and/or training the Project Advisors of the Pacific CHILD program. What is your role in preparing them to provide the trainers and the teachers with program content through the Mini-Institutes, Summer Institutes, and classroom demonstrations/observations? (Probe: Learning Team via Technology, weekly individual consultations)

2. How were the Pacific CHILD Project Advisors selected?
3. Please describe the training needs of the Pacific CHILD Project Advisors and trainers with whom you are working.

   Potential Probes:
   a. How much training experience do they have? How much teaching experience do they have? How much training of teachers experience do they have?
   b. What kinds of skills have you sought to develop with them? What skills have you emphasized as central to Pacific CHILD and/or working with trainers of teachers of ELLs?
   c. How would you assess their progress? What additional training do they need at this time?
   d. How much command do you feel the project advisors and the trainers have of its core concepts and components?
   e. How was it determined which islands would have island based staff. What affect, if any, does the lack of island based staff have on implementation?

Staff Needs for Professional Development

1. What kinds of training have the Project Advisors received in the last year?

2. What is your sense of the teachers' professional development needs in the language arts and ELL areas in CNMI, AS, and Hawaii?

3. How have you factored the teachers' needs into how you are working with the Project Advisors (probe for differences across entities)?

Training Institutes

1. What has your role been in developing the content, format and/or venue of the Summer and/or Mini-Institutes?

2. Please describe any training the Project Advisors and the trainers received in preparation for leading Summer and Mini-Institutes?

3. What your primary impressions of any of the Summer or Mini-Institutes you have attended? Please indicate the date and location of each Institute that you have attended and please describe your impressions of each of the following aspects of the Institute:

    a. Format
    b. Relevance/usefulness of content
    c. Role of trainers and other speakers
    d. Engagement of participants

4. How, if at all, have the institutes changed over the course of implementation? What are some of the reasons for these modifications? (Probe: school environment which affected the implementation of the mini institutes? For example, release time for teachers to observe demonstration lessons. How were the institutes modified in Hawaii in 2008-2009 based on your experiences in AS and CNMI in 2007-2008? Did the content change?)

5. Are there specific ways in which you think the format, content, or any other aspects of the institutes *should* change or be modified going forward to render them more effective?

Structured Learning Teams

1. Have you prepared or worked with the Project Advisors or trainers to help them ensure successful SLTs? Is so, how?

2. What is your understanding of the progress of the SLTs, i.e. **if they are meeting**, if teachers are collaborating and using them effectively, if they have helped teachers realize the goals of the program? Have the SLTs been able to reiterate, clarify and extend the major concepts that constitute Pacific CHILD and helped teachers apply them in the classroom? Have they been able to provide the teachers with support and motivation?

3. What is your impression of the degree to which SLTs represent a 'burden' for the teachers?

4. How, if at all, have the SLTs changed over the course of implementation? What are some of the reasons for these modifications?

5. Are there specific ways in which you think the format, content, or any other aspects of the SLTs should change or be modified going forward to render them more effective?

PREL Demonstrations and Classroom Observations

1. What is your understanding of the PREL demonstrations and classroom observations, i.e. How are they helping teachers realize the goals of the program? How are teachers responding to the pre and post conferences for the demonstrations and the PREL classroom observations?

2. Were there any additional modifications made to the professional development activities during the 2008-2009 school year? (probe: pre-observations, observations, post observations, demonstrations, etc)

Recruitment and Program Manuals

1. What was your recruitment strategy for Year 1 (2008-2009) in Hawaii? How did PREL choose the schools to be included in the study sample?

2. What kinds of incentives did you offer to teachers, schools and districts to encourage participation?

3. How were the Pacific CHILD manuals developed for Year 1 in Hawaii and Year 2 in AS and CNMI?

4. How and why was (what were the processes involved in) the decision made to revise the manuals for Year 2? How will they be revised for Year 2?

Application of the Pacific CHILD Components

1. At this point in the program, how much command do you feel the trainers have of Pacific CHILD's primary or core concepts and components? Probe for differences between the entities.

2. How much command do you feel the teachers have of its core concepts and components? Are there any particular components teachers have difficulty applying in the classroom?

3. What is your sense of the degree to which the teachers have been able *to apply* Pacific CHILD's core components in the classroom (with their existing curriculum)? What have been some of the gaps? Major successes with respect to classroom application/implementation?

4. In your opinion, how does Pacific CHILD fit into existing reading programs? (Probe: potential conflicts with other professional development activities, standards, benchmarks, school obligations, curriculum (informational vs narrative text) etc)

Initial Program Satisfaction and Impact

1. What is your sense of Pacific CHILD's Project Advisors' and trainers' level of satisfaction with the program at this point? What are some specific strengths and weaknesses that they have identified?

2. What is your sense of teachers' level of satisfaction with Pacific CHILD at this point? Specific strengths and weaknesses that they see in the program? Probe for SLTs, demos, observations, pre and post conferences.

3. If you think about program effectiveness, what kinds of impacts do you think Pacific CHILD is having on trainers, teachers and/or students?
    a. What, if any, impact has Pacific CHILD had on trainers and teachers knowledge, perceptions and attitudes toward ELL students?
    b. How has Pacific CHILD contributed to or supported teacher knowledge, skills, and/or classroom practices related to reading comprehension?
    c. How has it contributed to improved student achievement in reading comprehension?

4. Can you identify any specific instances where you have observed Pacific CHILD having an impact? (Probe: on trainers, teachers or students) Please describe.

School and District Support

1. Have you or the Project Advisors had much interaction with the schools and districts during the implementation of Pacific CHILD? If so, what kinds of interactions? If not, why not?

2. What is your sense of the degree to which the school and district administrators favor and support Pacific CHILD? If supportive, how? If not supportive, why not?

3. What is your sense of the level of support the teachers have received from their school and district administrators to participate in Pacific CHILD? Have there been some specific ways in which they have received support? Have there been specific barriers coming from the schools and/or districts that the teachers have experienced?

Overall Experience with Pacific CHILD

1. During program implementation in 2008-2009, what were some of the entity specific modifications or adaptations that were made to Pacific CHILD? Probe for specific examples.

2. Have you made any additional modifications to program implementation for Hawaii (or for Year 2 in CNMI and/or AS based on your experiences in Year 1)?

3. What have been some of the greatest obstacles to program implementation in each site?

4. What have been some of the greatest successes with respect to program implementation at each site?

5. Given the challenges and strengths you have identified, are you thinking of implementing any further changes to Pacific CHILD either at this point (mid-course corrections for Year 2in HI) or for future implementations of the program?

THANK YOU FOR YOUR TIME AND PARTICIPATION

L.7 Pacific CHILD interview guide: Project advisors

Date _____

Interviewer Name _____

Respondent Title _____

Interviewer Instructions

Where respondent is unfamiliar with the topic of a question, please skip or tailor to respondents specific role in Pacific CHILD, i.e. trainers of trainers vs. administrators of other types. Remember that different persons have different kinds of knowledge about Pacific CHILD, its purpose, implementation, and impact and therefore need to be encouraged to answer in a manner commensurate with their role.

Opening Script

Thank you for meeting with us. The purpose of our discussion today is to learn more about your experiences with Pacific CHILD. Your point of view as a Pacific CHILD project advisor is extremely valuable to us. We are especially interested in learning more about how you feel the professional development is going so far, and your impressions of teachers' progress in implementing the core elements of Pacific CHILD to work more effectively with ELLs.

We want today to focus on the implementation of Pacific CHILD during the 2008-2009 school year in American Samoa and CNMI (Year 2) and Hawaii (Year 1)

All of your answers to our questions today are entirely voluntary. We will maintain strict confidentiality for any answers you provide and your name will not be associated with any of your answers. Instead, your answers will be combined with those of others in our analyses and reporting, without your name attached to them, to help create a portrait of Pacific CHILD implementation.

We will record our conversation today for the purposes of note-taking only. All recordings and notes will be kept in secured, locked facilities for the duration of this project and destroyed upon project completion.

Do you have any questions before we begin?

Background

1. Please briefly tell me about your background and experience prior to working with PREL and/or on Pacific CHILD. (*Probe: years of teaching experience if any, experience with providing support to ELL teachers, experience with professional development programs)*

2. How long have you been with PREL? How long have you worked on the Pacific CHILD program?

3. Please describe your role as a Project Advisor in Pacific CHILD. (Probe: program design, manual design, manual revision, school recruitment, summer institutes, mini-institutes, demonstrations/observations, SLTs, school liaisons, etc.).

4. [*For those who identify a role in program design, please follow up with this question*:] Please describe your role in developing the content of the Pacific CHILD itself. How did you determine program content? What were your goals for the program? What did you hope the program would achieve?

5. What do you believe are the most critical components of Pacific CHILD? What do you think are its core elements as well as its most crucial components (greatest strengths)? Which components or aspects of the program do you feel should not be modified or adapted for a particular island or school?

6. How would you describe the role of the trainers in the Pacific CHILD program?

Training on Providing Professional Development

1. Did you receive any training on your role as a 'trainer-of-trainers' from PREL for the 2008-2009 implementation of Pacific CHILD? If so, what kinds of training did you receive?

2. What kind of on-going support or training do you receive from PREL or other sources as a 'trainer-of-trainers'?

Recruitment and Program Manuals

1. Were you involved in school recruitment? If so, do you know what the recruitment strategy for Year 2 (2008-2009) in Hawaii? How did PREL choose the schools to be included in the study sample?

2. What kinds of incentives did PREL offer the teachers, schools and districts to encourage participation?

3. Were you involved in developing or revising the Pacific CHILD manuals? If so, how were the Pacific CHILD manuals developed for Year 2?

4. How and why was (what were the processes involved in) the decision made to revise the manuals for Year 2? How will they be revised for Year 2? How has the Pacific CHILD content changed or modified for the 2008-2009 year (including the teacher and professional development manuals)?

Teachers' and Trainers' Need for Professional Development

1. What is your sense of the teachers' professional development needs in the language arts and ELL areas in CNMI, AS, and Hawaii?

2. How have you factored the teachers' needs into how you are working with the trainers of the teachers (probe for differences across entities)?

3. How were trainers selected? What kinds of training have the trainers needed? How have you worked to provide them with this training?(Probes: How much training experience do they have? How much teaching experience do they have?)

4. How was it determined which islands would have island based staff ? What affect, if any, does the lack of island based staff have on implementation?

Training Institutes

1. What has your role been in developing the content, format and/or venue of the Summer and/or Mini-Institutes?

2. How, if at all, have you trained the trainers to provide the institutes?

3. Have you provided any of the training directly to the teachers? If so, please explain your role.

4. What your primary impressions of any of the Summer or Mini-Institutes you have attended? Please indicate the date and location of each Institute that you have attended and please describe your impressions of each of the following aspects of the Institute:

    a. Format
    b. Relevance/usefulness of content
    c. Role of trainers and other speakers
    d. Engagement of participants

5. Please describe your impressions of the role and impact of the institutes.

6. How, if at all, have the institutes changed over the course of implementation? What are some of the reasons for these modifications? (Probe: school environment which affected the implementation of the mini institutes? For example, release time for teachers to observe demonstration lessons. How were the institutes modified in Hawaii in 2008-2009 based on your experiences in AS and CNMI in 2007-2008? Did the content change?)

7. Are there specific ways in which you think the format, content, or any other aspects of the institutes *should* change or be modified going forward to render them more effective?

Structured Learning Teams

1. How have you prepared for or worked with the trainers and the teachers to help them conduct successful SLTs?

2. Please describe your impressions of any of the SLTs you have observed or facilitated? Teacher level of engagement? Usefulness/importance to the PD? Effectiveness in reiterating and animating major concepts?

3. What is your understanding of the progress of the SLTs during the 2008-2009 year, **i.e., if they are meeting**, if teachers are collaborating and using them effectively, if they have helped teachers realize the goals of the PD, if they have provided support and motivation to the teachers?

4. What is your impression of the degree to which SLTs represent a 'burden' for the teachers?

8. How, if at all, have the SLTs changed over the course of implementation? What are some of the reasons for these modifications?

5. Are there specific ways in which you think the format, content, or any other aspects of the SLTs should change or be modified going forward to render them more effective?

PREL Demonstrations and Classroom Observations

6. What is your understanding of the PREL demonstrations and classroom observations, i.e. How are they helping teachers realize the goals of the program? How are teachers receiving pre and post conferences for the demonstrations and the PREL classroom observations?

7. Were there any additional modifications made to the professional development activities during the 2008-2009 school year? (probe: pre-observations, observations, post observations, demonstrations, etc)

Application of the Pacific CHILD Components

1. At this point in the program, how much command do you feel the **trainers** have of Pacific CHILD's primary or core concepts and components?

2. How much command do you feel the **teachers** have of its core concepts and components?

3. What is your sense of the degree to which the teachers have been able *to apply* Pacific CHILD's core components in the classroom (with their existing curriculum)? What have been some of the gaps? Major successes with respect to classroom application/implementation?

4. Are there any particular Pacific CHILD components or activities that the teachers are resistant to? If so, how have you or the trainers addressed this resistance?

5. In your opinion, how does Pacific CHILD fit into existing reading programs? (Probe: potential conflicts with other professional development activities, standards, benchmarks, curriculum (informational vs narrative text) etc)

Initial Program Satisfaction and Impact

1. What is your sense of Pacific CHILD's **trainers'** level of satisfaction with the program at this point? What are some specific strengths and weaknesses that they have identified?

2. What is your sense of **teachers'** level of satisfaction with Pacific CHILD at this point? Specific strengths and weaknesses that they see in the program? Probe for SLTs, demos, observations, pre and post conferences.

3. If you think about program effectiveness, what kinds of impacts do you think Pacific CHILD is having on trainers, teachers and/or students?

    a. On knowledge, perceptions and attitudes toward ELL students?
    b. On supporting teacher knowledge, skills, and/or classroom practices related to reading comprehension?
    c. On improved student achievement in reading comprehension?

School and District Support

4. Have you had much interaction with the schools and districts during the implementation of Pacific CHILD? If so, what kinds of interactions? If not, why not?

5. What is your sense of the degree to which the school and district administrators favor and support Pacific CHILD? How much support have the teachers received for their participation in Pacific CHILD? Please provide specific examples of support for and/or barriers to participation for the teachers.

Overall Experience with Pacific CHILD

6. During program implementation in 2008-2009 what have been some of the entity specific modifications or adaptations that were made to Pacific CHILD? Have you made any modifications to program implementation in Hawaii Year 2

7. What have been some of the greatest obstacles to program implementation in each site?

8. What have been some of the greatest successes with respect to program implementation at each site?

9. Do you feel that Pacific CHILD is helping the teachers better meet the needs of their students, especially ELLs? If so, how? If not, why not?

10. Given the challenges and strengths you have identified, please make any suggestions you have for changes to Pacific CHILD either at this point (mid-course corrections) or for future implementations of the program.

THANK YOU FOR YOUR TIME AND PARTICIPATION

L.8 Pacific CHILD interview guide: Trainers

Date _____

Interviewer Name _____

Respondent Title _____

Interviewer Instructions

Where respondent is unfamiliar with the topic of a question, please skip or tailor to respondents specific role in Pacific CHILD, i.e., trainers of trainers vs. administrators of other types. Remember that different persons have different kinds of knowledge about Pacific CHILD, its purpose, implementation, and impact and therefore need to be encouraged to answer in a manner commensurate with their role.

Opening Script

Thank you for meeting with us. The purpose of our discussion today is to learn more about your experiences with Pacific CHILD. Your point of view as a Pacific CHILD trainer is extremely valuable to us.

We are especially interested in learning more about how you feel the training and coaching sessions are going so far, and how you feel about the teachers' progress in implementing the core elements of Pacific CHILD in the classroom.

Today we would like to focus on the implementation of Pacific CHILD during the 2008-2009 school year in American Samoa and CNMI (Year 2) and Hawaii (Year 1)

All of your answers to our questions today are entirely voluntary. We will maintain strict confidentiality for any answers you provide and your name will not be associated with any of your answers. Instead, your answers will be combined with those of others in our analyses and reporting, without your name attached to them, to help create a portrait of Pacific CHILD implementation.

We will record our conversation today for the purposes of note-taking only. All recordings and notes will be kept in secured, locked facilities for the duration of this project and destroyed upon project completion.

Do you have any questions before we begin?

Background

1. Please briefly tell me about your background and experience prior to working with PREL and/or on Pacific CHILD. (*probe: years of teaching experience if any, experience with providing support to ELL teachers, length of experience with Pacific CHILD*)

2. How long have you been with PREL? How long have you worked on the Pacific CHILD program?

3. Please describe your role in Pacific CHILD. (Probe: program design, manual design, manual revision, school recruitment, summer institutes, mini-institutes, demonstrations/observations, SLTs, school liaisons, etc.).

4. *For those who identify a role in program design, please follow up with this question*: Please describe your role in developing the content of the Pacific CHILD itself. How did you determine program content? What were your goals for the program? What did you hope the program would achieve?

5. How has the Pacific CHILD content changed or modified for the 2008-2009 year (including the teacher and professional development manuals)?

6. What do you believe are the most critical components of Pacific CHILD? What do you think are its core elements as well as its most crucial components (greatest strengths)? Which components or aspects of the program do you feel should not be modified or adapted for a particular island or school?

Training on Providing Professional Development

1. Did you receive any training on your role as a 'trainer' from PREL for the 2008-2009 implementation of Pacific CHILD? If so, what kinds of training did you receive?

2. What kind of on-going support or training do you receive from PREL or other sources as a 'trainer-of-trainers'?

Teachers' Need for Professional Development

1. What is your sense of the teachers' professional development needs in the language arts and ELL areas?

2. How have you factored these needs into how you are working with the teachers?

Training Institutes

1. What has your role been in developing the content, format and/or venue of the Summer and/or Mini-Institutes?

2. Please describe your role in training the teachers directly at these institutes?

3. What your primary impressions of any of the Summer or Mini-Institutes you have participated in or attended? Please indicate the date and location of each Institute that you have attended and please describe your impressions of each of the following aspects of the Institute:
    a. Format
    b. Relevance/usefulness of content
    c. Role of trainers and other speakers
    d. Engagement of participants

4. In addition to providing training on the content of Pacific CHILD, how much support and motivation do you think the institutes have given the teachers?

5. What is your impression of how, if at all, have the institutes changed over the course of implementation? What are some of the reasons for these modifications and changes? (Probe: school environment which affected the implementation of the mini institutes? For example, release time for teachers to observe demonstration lessons. How were the institutes modified in Hawaii in 2008-2009 based on your experiences in AS and CNMI in 2007-2008? Did the content change?)

6. Are there specific ways in which you think the format, content, or any other aspects of the institutes *should* change or be modified going forward to render them more effective?

Structured Learning Teams

1. How have you prepared or worked with the teachers to help them conduct successful SLTs?

2. What were your some of your primary observations and impressions of the SLTs you have facilitated or observed? Teacher level of engagement? Usefulness/importance to the PD? Effectiveness in reiterating and animating major concepts? Ability to provide support for and motivation to teachers?

3. How much content do you think the SLTs have provided? Have they been able to reiterate, clarify and extend the major concepts that constitute Pacific CHILD and helped teachers apply them in the classroom?

4. What is your impression of the degree to which SLTs represent a 'burden' for the teachers?

5. How, if at all, have the SLTs <u>changed</u> over the course of implementation? What are some of the reasons for these modifications?

6. Are there specific ways in which you think the format, content, or any other aspects of the SLTs <u>should change</u> or be modified going forward to render them more effective?

PREL Demonstrations and Classroom Observations

1. What is your understanding of the PREL demonstrations and classroom observations, i.e. How are they helping teachers realize the goals of the program? How are teachers responding to the pre and post conferences for the demonstrations and the PREL classroom observations?

2. Were there any additional modifications made to the professional development activities during the 2008-2009 school year? (probe: pre-observations, observations, post observations, demonstrations, etc)

Application of the Pacific CHILD Components

1. At this point in the program, how much command do you feel the teachers have of Pacific CHILD's primary or core concepts and components?

2. What is your sense of the degree the teachers have been able *to apply* Pacific CHILD's core components in the classroom? What have been some of the gaps? Major successes with respect to classroom application/implementation? Are there any particular components teachers have had difficulty with?

3. In your opinion, how does Pacific CHILD fit into existing reading programs? (Probe: potential conflicts with other professional development activities, standards, benchmarks, curriculum (informational versus narrative text) etc)

Initial Program Satisfaction and Impact

1. What is your sense of teachers' level of satisfaction with Pacific CHILD at this point? Specific strengths and weaknesses that they see in the program? (Probe: SLTs, demos, observations, pre and post conferences)

2. If you think about program effectiveness, what kinds of impacts do you think Pacific CHILD is having on teachers and/or students?
   a. On knowledge, perceptions and attitudes toward ELL students?
   b. On supporting teacher knowledge, skills, and/or classroom practices related to reading comprehension?
   c. On improved student achievement in reading comprehension?

School and District Support

1. Have you had much interaction with the schools and districts during the implementation of Pacific CHILD? If so, what kinds of interactions? If not, why not?

2. What is your sense of the degree to which the school and district administrators favor and support Pacific CHILD? How much support have the teachers received for their participation in Pacific CHILD? Please provide specific examples of support for and/or barriers to participation for the teachers.

Overall Experience with Pacific CHILD

1. What have been some of the greatest obstacles to implementing Pacific CHILD in CNMI or AS or Hawaii?

2. What have been some of the greatest successes with respect to program implementation at each site?

3. Do you feel that Pacific CHILD is helping the teachers better meet the needs of their students, especially ELLs? If so, how? If not, why not?

4. Given the challenges and strengths you have identified, do you have any suggestions for changes you think should be made to Pacific CHILD either at this point or for future implementations of the program.

THANK YOU FOR YOUR TIME AND PARTICIPATION

L.9 Teacher focus group discussion guide

Thank you for meeting with us today. The purpose of our visit is to learn more about your experiences with the Pacific CHILD professional development program. Your point of view as teacher participants is extremely valuable to us. We are especially interested in learning more about how you feel the training sessions are going so far, and how you feel about your progress in implementing key components of the Pacific CHILD in your classroom.

Please keep in mind:

- There are no right or wrong answers to the questions we have prepared. Our questions are designed to stimulate discussion about your experiences with Pacific CHILD.

- Your responses are confidential – we will not identify you or your school when we summarize what we have learned from all the participating teachers in our evaluation report. Your answers will not be shared with other teachers, principal, reading specialists, administrators, or any of the PREL staff.

- We will take notes during the discussion for our internal use only. We will not share these notes with anyone outside of the BPA evaluation team. Your comments will be kept anonymous and will not be traceable to you. Do you mind if we record the session just to help us with our notes? No one else will have access to the tape – it is just for us.

- Please remember to be respectful of everyone's opinion. You do not need to agree with each other or reach consensus.

- Your participation in this focus group is completely voluntary and you do not have to answer any questions that you do not want to answer.

A. BACKGROUND

1. Now, let's start by everyone sharing their name and the grade you are teaching..

2. Briefly explain what kind of professional development activities you have participated in over the past year? (content, format and time commitment)

B. APPLYING PACIFIC CHILD PRINCIPLES/COMPONENTS – FIT

1. How does Pacific CHILD fit into your existing reading program? How well do you feel you are able to implement Pacific CHILD strategies in your classroom?

2. What has been most useful and relevant?
    - Are there any elements of the PD that you find particularly relevant for you and for your students in your school? If so, what are they?
    - How do these elements help you to work more effectively with your students?
    - Do you use Pacific CHILD strategies outside of your language arts lessons? How?
    - How has what you have learned changed how you work with your students in your classroom?
    - If nothing has changed, why do you think this is?

3. What do you think of the cultural component of the program? To what extent Do you think it was successful in focusing on cultural context of the Pacific is helpful?

4. Now that you have had two years of experience with the Pacific CHILD model, how do you think it is relevant to addressing the needs of ELLs?
    - Have you found that Pacific CHILD has been able to support the specific needs of your ELL students? If so, how?
    - Since you started the professional development program have you changed anything you are doing in the classroom with ELLs? If so, what have you changed?
    - Do you feel more/less confident about meeting the needs of ELLs?

5. Do you find that Pacific CHILD helps you deal with students of different academic levels? If so, how?

6. How have your students (in general) responded to the Pacific CHILD activities?
    - Can you see any changes in your students now that you have introduced some ideas you have learned through Pacific Child into your classroom? What changes? Are they benefiting your students? Or distracting them from other goals you are trying to accomplish with them?
    - If no changes, why do you think nothing has changed?
    - Have you seen any specific improvements in your students' academic skills since you began using Pacific CHILD in your classroom, e.g., improved understanding of why they are learning what they are learning?
    - When you think about the entire study (2-year study, although only 1 year of involvement for many students) – do you expect we will be able to see changed in students' academic skills reflected in test scores? Why or why not?

C. IMPLEMENTATION CONTEXT

1. Are there any particular characteristics of you school that you think have affected the implementation of Pacific CHILD? (Has school adopted reading curriculum? Are ESL students pulled out? Does each teacher teach all subjects or do teachers team teach?)

2. Do you find Pacific CHILD and your other professional development programs are compatible? Complementary? Contradictory? Please describe how.

3. In addition to professional development activities, please describe any other obligations you may have at your school, e.g. staff meetings, committee meetings, official conferences or other obligations. Have you found that Pacific Child activities detract from other goals you are trying to achieve? If so, how?

4. Have you had access to sufficient materials, e.g., informational texts, instructional aids, supplies, books, needed equipment and other teaching materials, over the past year to be able to implement Pacific CHILD in your classroom? Has it been a burden to find the text you need?

5. How well do you feel your school and district level administrators support the Pacific CHILD program?

   - What kinds of supports have your principles, schools, and school districts given you for your participation in Pacific CHILD? Please provide specific examples.
   - Have your schools provided you with the resources you need to implement Pacific CHILD? Please provide specific examples.
   - Ability to leave classroom to observe demonstration lessons?
   - Has the school provided any additional time or scheduling flexibility to make it possible to fit Pacific CHILD into your schedule?
   - Are there additional supports or resources you would like your principles, schools, and school districts to provide in order to better support your participation in Pacific CHILD?

D. THE PROFESSIONAL DEVELOPMENT MODEL

Let's think a little about the summer institutes, mini institutes, SLTs, including PREL demo lessons, lesson observations by PREL, including pre and post conferences

Components

1. What were some of the most helpful components of Pacific CHILD that you tried to use this year? How were they helpful?

2. What were some of the least helpful or most challenging components of Pacific CHILD that you tried to use this year? Why were they challenging? What did you do in response to their level of difficulty?

   - If some components are not relevant, why do you think they are not relevant? Are the principles or ideas inappropriate or uninteresting? If so, how?

Institutes

1. What were the **most** helpful or useful aspects of the 2009 Summer Institute?
2. What were the least helpful or useful aspects of the 2009 Summer Institute?
3. What were the **most** helpful or useful aspects of the 2009-10 Mini Institutes?
4. What were the **least** helpful or useful aspects of the 2009-10 Mini Institutes?

SLTs

1. How often does PREL attend an SLT?
2. How often does the SLT meet without PREL?
   - How long?
   - What do you discuss?
   - Does everyone attend? (If they do not regularly participate, why? -- schedule conflict, lack of time, lack of resources, not interested, etc)
   - What are some of the benefits you feel you have received from the structured learning teams? How helpful are they?
   - Teacher collaboration and support – part of the school culture prior to Pacific CHILD, has participating in Pacific CHILD influenced that?

Coaching and Feedback

To what extent are you getting feedback and coaching from PREL on your teaching practices and how helpful is it?

- As you design your lessons
- Pre and post conferences (are these occurring? In person, by phone, or email?)
- During SLTs
- Ask questions and get info by email or on phone

Overall

What do you think of the overall format of the PD of a 2-week summer institute followed by 3 minis, demonstration lessons and PREL observation of teacher lessons?

- What works well?
- What doesn't and why?
- Any other ideas or suggestions about how the format of the PD and the different components could be improved?

E. CROSS-OVER ISSUES

1. Have you shared any of your Pacific CHILD materials or manuals with teachers who are not in the program? What did they think of them?

2. Have you discussed what you are learning in the program with colleagues who are not in the program? If so, what kinds of topics have you discussed?

3. Has this been helpful to you to talk or collaborate with teachers who are not in the program? Has this discussed changed how you are using what you have learned through Pacific CHILD? If so, how?

F. OVERALL SATISFACTION WITH PACIFIC CHILD

1. Looking back at your experience with Pacific CHILD so far, how would you describe your overall level of satisfaction with the program?

   Probes*: Do you feel you are making progress? Do you feel the program is helping you meet the needs of your students? What is the most helpful component of the Pacific CHILD Program so far?*

2. Do you have any other suggestions for improving the program?

THANK YOU

L.10 Professional development observation summary for Annual Institute Week 1

2009 Pacific CHILD Annual Institute

Professional Development Observation Summary Sheet: WEEK 1

Island:

To complete this form, please delete the text in the grayed out boxes, and insert your text in its place.

Dates of Observation: For example, June 22-26 (Days 1-5)

Location: For example, XX Elementary School library, XXX. Oahu

Name(s) of Observers(s):

Name(s) of Presenter(s):

Participant Information:

Number of teachers from each school, information on teacher attendance (absences, missed sessions, etc.):

Description of Room Arrangements:

Description of Grouping Assignments:

Grouped by school (SLTs), across schools (CLGs), etc.

Implementation/Fidelity to Design:

Compare actual versus planned in PD manual

Insert Timeline for each day observed

Sample timetable: Annual Institute Day 1 (content from "Activity as indicated in the PD manual" column removed for this report)

Time as indicated in the PD	Activity as indicated in the PD	Time as indicated on the agenda	Activity as indicated on the agenda	Actual Time	Actual Activity
8:00–8:30					
8:30–9:50					
10:00–11:30					
Lunch					
12:30–2:00					
2:00–3:00					
3:00–3:15					

<u>Summarize</u> the Use and Connections to the Six Pacific CHILD Components

1. Vocabulary: a) Word Knowledge b) Word Parts
2. Question Generation:
3. Text Structure: a) Sequence b) Compare and Contrast c) Cause and Effect d) Text Features
4. Differentiated Instruction:
5. Interactive Tasks:
6. Cognitively-rich Environment:

Facilitation and Pacing

(e.g what is your impression of the amount of information trainers were covering? Where they going too fast? Too slow? Were the topics clear? Provide specific information about each trainer)

Engagement

(e.g how would you assess teachers' level of engagemet during the lectures, tasks, and demos? What was your impression of teacher satisfaction w/the institute? Which areas did teachers appear to enjoy the most? Areas teachers struggled with?

Concerns voiced by teachers and how/if they were addressed:

Community Support:

Presence of principals, Dept of Ed., etc.)

Applicability/Appropriateness of Pacific CHILD:

Include teacher and PREL staff remarks about the applicibility of Pacific CHILD to existing language arts classrooms and curriculum, standards/benchmarks, student proficiency levels, and the cultural backgrounds of the teachers and their students

Reading Assignments/Homework:

Impressions:

Miscellaneous:

L.11 Professional development observation summary for Annual Institute Week 2

2009 Pacific CHILD Annual Institute

Professional Development Observation Summary Sheet: WEEK 2

Island:

To complete this form, please delete the text in the grayed out boxes, and insert your text in its place.

Dates of Observation: June 22-26 (Days 1-5)

Location: i.e., XXX Elementary School library, XXX, Oahu

Name(s) of Observers(s):

Name(s) of Presenter(s):

Participant Information:

Number of teachers from each school, information on teacher attendance (absences, missed sessions, etc.):

Description of Room Arrangements:

Air conditioned library with x, y, z.

Implementation/Fidelity to Design:

Compare actual versus planned in PD manual

Day	Planned Foci (PD manual)	Actual Foci
6	Compare and Contrast, Word Parts	
7	Compare and Contrast, Question Generation	
8	Cause and Effect, Words,	
9	Cause and Effect, Question Generation	
10	Sequence, Words, Word Parts	

Insert Timeline for each day observed

Sample timetable: Annual Institute Day 5 (content from "Activity as indicated in the PD manual" column removed for this report)

Time as indicated in the PD	Activity as indicated in the PD	Time as indicated on the agenda	Activity as indicated on the agenda	Actual Time	Actual Activity
8:00–8:30					
8:30–9:50					
10:00–11:30					
Lunch					
12:30–2:00					
2:00–3:00					
3:00–3:15					

Facilitation and Pacing:

(e.g what is your impression of the amount of information trainers were covering? Where they going too fast? Too slow? Were the topics clear? Provide specific information about each trainer)

Engagement:

(e.g how would you assess teachers' level of engagemet during the lectures, tasks, and demos? What was your impression of teacher satisfaction w/the institute? Which areas did teachers appear to enjoy the most? Areas teachers struggled with?

Concerns voiced by teachers and how/if they were addressed:

Community Support:

Presence of principals, Dept of Ed., etc.)

Applicability/Appropriateness of Pacific CHILD

Include teacher and PREL staff remarks about the applicibility of Pacific CHILD to existing language arts classrooms and curriculum, standards/benchmarks, student proficiency levels, and the cultural backgrounds of the teachers and their students

Reading Assignments/Homework:

Impressions:

Miscellaneous:

L.12 Professional development observation summary for Mini Institute full day

2009 Pacific CHILD Mini Institute FULL DAY template

Professional Development Observation Summary Sheet: Mini #__

Island:

To complete this form, please delete the text in the grayed out boxes, and insert your text in its place.

Dates of Observation: For example, September 12, 2009 (full day mini)

Location: For example, XXX Elementary School library, XXXX, Oahu

Name(s) of Observers(s):

Name(s) of Presenter(s):

Participant Information:

Number of teachers from each school, information on teacher attendance (absences, missed sessions, etc.):

Description of Room Arrangements:

Description of Grouping Assignments:

Grouped by school (SLTs), across schools (CLGs), etc.

Implementation/Fidelity to Design:

Compare actual versus planned in PD manual. Include a fidelity summary based on completed timetable

Insert Timeline for each day observed

Sample Mini Institute Timeline

Time as indicated in the PD	Activity as indicated in the PD	Time as indicated on the agenda	Activity as indicated on the agenda	Actual Time	Actual Activity
8:00–8:20					
8:20–9:50					
10:00–11:30					
Lunch					
12:30–2:30					
2:30–3:00					

Mini Institute Component Foci (planned and actual according to PD manual)

Simply list planned component foci and actual foci

<u>Summarize</u> the Use and Connections to the Six Pacific CHILD Components

1. Vocabulary: a) Word Knowledge b) Word Parts
2. Question Generation:
3. Text Structure: a) Sequence b) Compare and Contrast c) Cause and Effect d) Text Features
4. Differentiated Instruction:
5. Interactive Tasks:
6. Cognitively-rich Environment:

Facilitation and Pacing

(e.g what is your impression of the amount of information trainers were covering? Where they going too fast? Too slow? Were the topics clear? Provide specific information about each trainer)

Engagement:

(e.g how would you assess teachers' level of engagemet during the lectures, tasks, and demos? What was your impression of teacher satisfaction w/the institute? Which areas did teachers appear to enjoy the most? Areas teachers struggled with?

Concerns voiced by teachers and how/if they were addressed

Community Support:

Presence of principals, Dept of Ed., etc.)

Applicability/Appropriateness of Pacific CHILD:

Include teacher and PREL staff remarks about the applicibility of Pacific CHILD to existing language arts classrooms and curriculum, standards/benchmarks, student proficiency levels, and the cultural backgrounds of the teachers and their students

Reading Assignments/Homework:

Impressions:

Miscellaneous:

L.13 Professional development observation summary for Mini Institute school-based day

2010 Pacific CHILD Mini Institute-SCHOOL BASED DAY template

Professional Development Observation Summary Sheet: Mini #_

Island:

To complete this form, please delete the text in the grayed out boxes, and insert your text in its place.

Dates of Observation: September 14-18

Location:

Full day Mini Institute these teachers attended:

Name(s) of Observers(s):

Name(s) of Presenter(s):

Participant Information:

Number of teachers from each school, information on teacher attendance (absences, missed sessions, etc.):

Description of Room Arrangements:

Implementation/Fidelity to Design:

Compare actual versus planned in PD manual. Include a school based fidelity summary based on the completed timelines

Day	Planned Foci (PD manual)	Actual Foci
1		
2		
3		
4		

Insert Timeline for each day observed

Facilitation and Pacing:

(e.g what is your impression of the amount of information trainers were covering? Where they going too fast? Too slow? Were the topics clear? Provide specific information about each trainer)

Engagement:

(e.g how would you assess teachers' level of engagemet during the lectures, tasks, and demos? What was your impression of teacher satisfaction w/the institute? Which areas did teachers appear to enjoy the most? Areas teachers struggled with?

Concerns voiced by teachers and how/if they were addressed:

Community Support:

Presence of principals, Dept of Ed., etc.)

Applicability/Appropriateness of Pacific CHILD:

Include teacher and PREL staff remarks about the applicibility of Pacific CHILD to existing language arts classrooms and curriculum, standards/benchmarks, student proficiency levels, and the cultural backgrounds of the teachers and their students

Reading Assignments/Homework:

Impressions:

Miscellaneous:

L.14 Program staff lesson demonstration observation summary

Day: **Date:**

PREL presenter(s):
Pre: Conference:
Length:
Teachers Present: (e.g. all 6 school X teachers)

Did teachers receive a lesson plan? Yes/No
Did teachers complete an observation form during the demonstration? Yes/No

Pacific CHILD Components:

Planned Foci according to PD manual: List components

Pacific CHILD components planned (see lesson plan): List components

Actual Pacific CHILD components covered: Include components and brief description of lesson

Lesson Length:

Post conference:

Length:
Teachers Present:

Impressions:

Include your impressions of the demo and the pre and post conferences

Reference: Pacific CHILD Components

1. Vocabulary	Comprehension Strategies	Format of Instruction
- Word Knowledge - Word Parts	2. Question Generation 3. Text Structure - Sequence - Compare and Contrast - Cause and Effect - Text Features	4. Differentiated Instruction 5. Interactive Tasks 6. Cognitively-rich environment

L.15 Teacher lesson demonstration observation summary

Teacher (s) presenting lesson: include school name
PREL staff observer (s):
Pre: Conference:
Length:
Teachers Present:

Did teachers receive a lesson plan? Yes/No
Did teachers complete an observation form during the demonstration? Yes/No

Pacific CHILD Components:

Planned Foci according to PD manual: List components

Pacific CHILD components planned (see lesson plan): List components

Actual Pacific CHILD components covered: Include components and brief description of lesson

Lesson Length:

Post conference:

Length:
Teachers Present:
Concerns:

Impressions:

Include your impressions of the demo and the pre and post conferences

Reference: Pacific CHILD Components

1. Vocabulary	**Comprehension Strategies**	**Format of Instruction**
- Word Knowledge - Word Parts	2. Question Generation 3. Text Structure - Sequence - Compare and Contrast - Cause and Effect - Text Features	4. Differentiated Instruction 5. Interactive Tasks 6. Cognitively-rich environment

L.16 Structured Learning Team observation (SLT)

Day : **Date**

PREL staff: include school name
Teachers Present: (e.g. all 6 school XX teachers)
Brief summary of SLT:

Pacific CHILD component covered:

Planned Foci according to PD manual: List components

Actual Pacific CHILD components covered: Include components and brief description of lesson

Concerns voiced by teachers or PREL staff::

Include your impressions of the demo and the pre and post conferences

Impressions:

Include your impressions of the demo and the pre and post conferences

Reference: Pacific CHILD Components

1. Vocabulary	**Comprehension Strategies**	**Format of Instruction**
- Word Knowledge - Word Parts	2. Question Generation 3. Text Structure - Sequence - Compare and Contrast - Cause and Effect - Text Features	4. Differentiated Instruction 5. Interactive Tasks 6. Cognitively-rich environment

References

Afflerbach, P. (2007). *Understanding and using reading assessment, K-12*. Newark, DE: International Reading Association.

Allison, P. D. (2001). *Missing data*. Thousand Oaks, CA: Sage.

American Samoa Department of Education. (2007a). *Certification program and reclassification for teachers*. Pago Pago, American Samoa: Author.

American Samoa Department of Education (2007b). *Fourth grade student Stanford Achievement Test 10 results* [Data file]. Pago Pago, American Samoa: Author.

Anders, P., Hoffmann, J., & Duffy, G. (2000). Teaching teachers to teach reading: Paradigm shifts, persistent problems, and challenges. In M. Kamil, P. Mosental, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3, pp. 721–744). Mahwah, NJ: Lawrence Erlbaum Associates.

Armbruster, B. B., Anderson, T. H., & Ostertag, J. (1987). Does text structure/summarization instruction facilitate learning from expository text? *Reading Research Quarterly*, *22*(3), 331–346.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a new and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, *57*(1), 1289–1300.

Bloom, H. S., Bos, J., & Lee, S. (1999). Using cluster random assignment to measure program impacts: Statistical implications for evaluation of education programs. *Evaluation Review*, *23*(4), 445–469.

Bloom, H. S., Richburg-Hayes, L., & Black, A. R. (2005). Using covariates to improve precision: empirical guidance for studies that randomize schools to measure the impacts of educational interventions (MDRC Working Papers on Research Methodology). New York: MDRC.

Bloom, H. S., Richburg-Hayes, L., & Black, A. R. (2007). Using covariates to improve precision for studies that randomize schools to evaluate educational interventions. *Educational Evaluation and Policy Analysis*, *29*(1), 30–59.

Birman, B., Le Floch, K. C., Klekotka, A., Ludwig, M., Taylor, J., Walters, K., et al. (2007). *State and local implementation of the No Child Left Behind Act: Vol. 2. Teacher quality under NCLB: Interim report.* Washington, DC: U.S. Department of Education, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service.

Burger, D., Mauricio, R., & Ryan, J. (2007). *English language proficiency assessment in the Pacific Region* (Issues & Answers Report, REL 2007–014). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Pacific.

Campbell, M. K., Elbourne, D. R., & Altman, G. A. (2004). CONSORT statement: Extension to cluster randomised trials. *British Medical Journal*, *328*(7441), 702–708.

Center for Research on Education, Diversity, and Excellence (CREDE). (2002). *The five standards for effective pedagogy*. Retrieved July 27, 2010, from http://crede.berkeley.edu/research/crede/standards.html.

Central Intelligence Agency. (2011). *The world factbook 2011.* Retrieved December 1, 2011, from https://www.cia.gov/library/publications/the-world-factbook/index.html.

Chesswas, R., Keir, S., Leung, E., & Terada, W. (2005). *Evaluation of the Pacific CHILD professional development program*. Honolulu, HI: Pacific Resources for Education and Learning.

Cole, D. C. (1992). The effects of a one-year staff develop-ment program on the achievement of test scores of fourth-grade students. *Dissertation Abstracts International*, *53*(06), 1792A. (UMI No. 9232258)

Cooper, H. M., Hedges, L. V., & Valentine, J. (Eds.) (2009). *The Handbook of Research Synthesis and meta-analysis* (2nd ed.). New York: The Russell Sage Foundation.

CNMI Public School System. (2007). *Fifth grade student Stanford Achievement 10 results* [Data file]. Saipan, Commonwealth of the Northern Mariana Islands: Author.

CNMI Public School System. (n.d.). *Jobs vacancies announcement*. Retrieved July 29, 2010,from www.cnmipss.org/state-level-programs/human-resources-office/job-vacancies-announcement/.

CTB/McGraw-Hill. (2003). *TerraNova, the second edition technical report*. Monterey, CA: Author.

Darling-Hammond, L., & McLaughlin, M. W. (1995). Policies that support professional development in an era of reform. *Phi Delta Kappa*, *76*(8), 597–604.

Dickson, S. V., Simmons, D. C., & Kame'enui, E. J. (1995). *Text organization and its relation to reading comprehension: A synthesis of the research*. Eugene, OR: National Center to Improve the Tools of Educators.

Duffy, G. G., Roehler, L. R., Meloth, M. S., Vavrus, L. G., Book, C., Putnam, J., et al. (1986). The relationship between explicit verbal explanations during reading skill instruction and student awareness and achievement: A study of reading teacher effects. *Reading Research Quarterly*, *21*(3), 237–252.

Duke, N. (2004). The case for informational text. *Educational Leadership*, *61*(6), 40–44.

Ebel, R. L., & Frisbie, D. A. (1991). *Essentials of educational measurement* (5th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Echevarria, J., & Short, D. (2004). Using multiple perspectives in observations of diverse classrooms: The Sheltered Instruction Observation Protocol (SIOP). In H. Waxman, R. Tharp, & S. Hilberg (Eds.). *Observational research in U.S. classrooms: New approaches for understanding cultural and linguistic diversity* (pp. 21–43). Boston: Cambridge University Press.

Echevarria, J., & Short, D. (n.d.). *The Sheltered Instruction Observation Protocol (SIOP)*. Retrieved July 18, 2010, from www.siopinstitute.net/media/pdfs/sioppaper.pdf.

Echevarria, J., Vogt, M., & Short, D. (2003). *Making content comprehensible for English learners* (2nd ed.). Boston: Allyn & Bacon.

Echevarria, J., Vogt, M., & Short, D. (2007). *Making content comprehensible for English learners* (3rd ed.). Boston: Allyn & Bacon.

Fullan, M. (2001). *The new meaning of educational change* (3rd ed.). New York: Teachers College Press.

Garet, M. S., Cronen, S., Eaton, M., Kurki, A., Ludwig, M., Jones, W., et al. (2008). *The impact of two professional development interventions on early reading instruction*. (NCEE 2008–4030). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

Genesee, F. (1994). *Integrating language and content: Lessons from immersion*. Washington, DC: National Center for Research on Cultural Diversity and Second Language Learning, Center for Applied Linguistics.

Gersten, R., Dimino, J. & Jayanthi, M. (2010). Teacher study group: Impact of the professional development model on reading instruction and student outcomes in first grade classrooms. *American Educational Research Journal*, *47*(3), 694–739.

Gibbons, P. (2006). *Bridging discourses in the ESL classroom*. London: Continuum.

Guarino, A. J., Echevarria, J., Short, D., Schick, J. E., Forbes, S., & Rueda, R. (2001). The *Sheltered Instruction Observation Protocol*. *Journal of Research Education*, *11*(1), 138–140.

Grigg, W. S., Daane, M. C., Jin, Y., & Campbell, J. R. (2003). *The nation's report card: Reading 2002* (NCES 2003–521).Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.

Haertel, E. (1986). One valid use of student performance measures for teacher evaluation. *Education evaluation and policy analysis, 8*(1), 45–60.

Hawaii Department of Education. (2007). *2007: Superintendent's 18th annual report*. Honolulu, HI: State of Hawai'i Department of Education

Hawaii Department of Education. (n.d.). *Furloughs and Hawai'i 's public schools*. Retrieved April 1, 2011, from http://doe.k12.hi.us/news/furlough/index.htm.

Hawaii Teacher Standards Board. (n.d.). *Types of licenses*. Retrieved July 22, 2010, from http://htsb.org/html/details/licensing/types.html.

Heine, H., & Emesiochl, M. (2007). *Preparing and licensing high quality teachers* (Issues & Answers Report, REL 2007–031). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Pacific.

Hiebert, E., & Kamil, M. (Eds). (2005). *Teaching and learning vocabulary*. Mahwah, NJ: Lawrence Erlbaum Associates.

Hunkin-Finau, S. S. (2007). *NCLB and education in the Pacific: A dire need for a balanced perspective—What do we do now?* Honolulu, HI: Pacific Resources for Education and Learning.

Jerry, L., & Lutkus, A. (2003). *The nation's report card: State reading 2002, report for American Samoa* (NCES 2003–526 AS). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.

Johnson, D. W., & Johnson, R. (1999). What makes cooperative learning work. In D. Kluge, S. McGuire, D. Johnson, & R. Johnson (Eds.), *Cooperative learning* (pp. 23–36). Tokyo: Japan Association for Language Teaching.

Joyce, B., & Showers, B. (1996). Staff development as a comprehensive service organization. *Journal of Staff Development*, *17*(1), 2–6.

Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices* (2nd ed.). New York: Springer.

Koretz, D. (2008). *Measuring up: What educational testing really tells us*. Cambridge, MA: Harvard University Press.

Learning First Alliance. (2000). *Every child reading: A professional development guide.* Washington, DC: Author.

Lee, J., Grigg, W., & Donahue, P. (2007). *The nation's report card: Reading 2007* (NCES 2007–496). Washington, D.C: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*(2), 149–174.

McCutchen, D., Abbott, R. D., Green, L. B., Beretvas, S. N., Cox, S., Potter, N. S., et al. (2002). Beginning literacy: Links among teacher knowledge, teacher practice, and student learning. *Journal of Learning Disabilities*, *35*(1), 69–86.

McGill-Franzen, A., Allington, R. L., Yokoi, L., & Brooks, G. (1999). Putting books in the classroom seems necessary but not sufficient. *Journal of Reading Research*, *93*(2), 67–74.

Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York: Macmillan.

National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction.* Bethesda, MD: National Institute of Child Health and Human Development.

Pacific Resources for Education and Learning (2007). *Pacific CHILD teacher's manual*. Honolulu, HI: Author.

Pearson. (2004). *Technical data report*. Bloomington, MN: Author.

Puma, M. J., Olsen, R. B., Bell, S. H., & Price, C. (2009). *What to do when data are missing in group randomized controlled trials* (NCEE 2009–0049). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

RAND Reading Study Group. (2002). *Reading for understanding: Towards an R&D program in reading comprehension.* Santa Monica, CA: RAND.

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

Rosenshine, B., Meister, C., & Chapman, S. (1996). Teaching students to generate questions: A review of the intervention studies. *Review of Educational Research*, *66*(2), 181–221.

Rueda, R. (1998). *Standards for professional development: A sociocultural perspective*. Santa Cruz, CA: Center for Research on Education, Diversity, and Excellence.

Schochet, P. Z. (2005). *Statistical power for random assignment evaluations of education programs.* Princeton, NJ: Mathematica Policy Research.

Schochet, P. Z. (2008). *The late pretest problem in randomized control trials of education interventions* (NCEE 2009–4033). Washington, DC: U.S. Department of Education, Institutes for Education Sciences, National Center for Education Evaluation and Regional Assistance.

Schochet, P. Z. (2009). *Technical methods report: The estimation of average treatment effects for clustered RCTs of education interventions* (NCEE 2009–0061rev). Washington, DC: U.S. Department of Education, Institutes for Education Sciences, National Center for Education Evaluation and Regional Assistance.

Sloan, H. A. (1993). Direct instruction in fourth and fifth grade classrooms. *Dissertation Abstracts International*, *54*(08), 2837A. (UMI No. 9334424)

Spybrook, J., Raudenbush, S. W., Congdon, R., & Martínez, A. (2009). *Optimal design for longitudinal and multilevel research: Documentation for the "Optimal Design" software: A technical document for Optimal Design version 2.0.* New York, NY: William T. Grant Foundation

Strickland, D., & Kamil, M. (Eds). (2004). *Professional development for teaching reading*. Norwood, MA: Christopher-Gordon.

Teacher reclassification bill. S.B. 30–78. 13th legislature of American Samoa. (2008) (enacted).

Tharp, R. G. (1997). *From at-risk to excellence: Research, theory, and principles for practice*. Santa Cruz, CA: Center for Research on Education, Diversity, and Excellence.

Tharp, R. G., Estrada, P., Dalton, S. S., & Yamauchi, L. A. (2000). *Teaching transformed: Achieving excellence, fairness, inclusion and harmony*. Boulder, CO: Westview Press.

Tienken, C. H. (2003). The effect of staff development in the use of scoring rubrics and reflective questioning strategies on fourth-grade students' narrative writing performance. *Dissertation Abstracts International*, *64*(02), 388A. (UMI No. 3081032)

U.S. Census Bureau. (2011a). *Annual estimates of the population for the United States, regions, states, and Puerto Rico: April 1, 2010 to July 1, 2011 (NST-EST2011-01).* Retrieved December 20, 2011 from http://www.census.gov/popest/data/state/totals/2011/index.html.

U.S. Census Bureau. (2011b). *2010 state area measurements and internal point coordinates*. Retrieved December 20, 2011 from http://www.census.gov/geo/www/2010census/statearea_intpt.html

U.S. Department of Education. (2005). *Improving teacher quality in U.S. school districts: Districts' use of Title II, Part A, funds in 2002–2003*. Washington, DC: Office of the Undersecretary, Policy and Program Studies.

U.S. Department of Education. (2007). *State and local implementation of the No Child Left Behind Act, Volume I—Teacher quality under NCLB: Interim report*. Washington, DC: Office of Planning, Evaluation, and Policy Development, Policy and Program Studies.

U.S. Department of Education (2009). *The Secretary's sixth annual report on teacher quality: A highly qualified teacher in every classroom.* Washington, DC: Office of Postsecondary Education.

U.S. Department of Education, Institutes of Education Sciences. (2008). *What Works Clearinghouse Procedures and Standards Handbook, Version 2*. Washington, DC: Author.

U.S. Department of Education, National Center for Education Statistics. (2007). Common Core of Data. Local Education Agency Universe Survey, 2006–07. Retrieved July 29, 2010, from http://nces.ed.gov/ccd/.

U.S. Department of Education, National Center for Education Statistics. (2008). Common Core of Data. Local Education Agency Universe Survey, 2007–08. Retrieved July 29, 2010, from http://nces.ed.gov/ccd/.

U.S. Department of Education, National Center for Education Statistics. (2010a). Public Elementary/Secondary School Universe Survey 2006/07. Retrieved July 29, 2011, from http://nces.ed.gov/ccd/.

U.S. Department of Education, National Center for Education Statistics. (2010b). Public Elementary/Secondary School Universe Survey 2007/08. Retrieved July 29, 2011, from http://nces.ed.gov/ccd/.

Wayne, A. J., Yoon, K. S., Zhu, P., Cronen, S., & Garet, M. S. (2008). Experimenting with teacher professional development: Motives and methods. *Educational Researcher, 37*(8), 469–479.

West, B. T, Welch, K. B., & Galecki, A. T. (2007). *Linear mixed models: A practical guide using statistical software*. Boca Raton, FL: Chapman & Hall/CRC.

Yoon, K. S., Duncan, T., Lee, S. W.-Y., Scarloss, B., and Shapley, K. 2007. *Reviewing the evidence on how teacher professional development affects student achievement* (Issues & Answers Report, REL 2007–033). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest.

Zieky, M. (2003). *A DIF primer*. Princeton, NJ: Educational Testing Services.

DEPARTMENT OF EDUCATION
UNITED STATES OF AMERICA