

## B. Collections of Information Employing Statistical Methods

### 1. Respondent Universe and Sampling Methods

#### Program Universe

The potential universe of programs includes all child care homes and child care centers in the geographic areas designated below who have the following characteristics:

- Serve infants and toddlers under the age of three (at least three children between the ages of three months and twenty-four months at the time of entry into the study)
- Have been in operation for at least one year.
- Have not previously participated in the Program for Infant Toddler Care.
- English or Spanish is the primary language spoken in the classroom.

The greater Los Angeles, Phoenix, and Tucson metro areas have large numbers of child care providers who could benefit from the PITC program. Exhibit 9 shows the numbers of providers in Riverside, Orange, San Bernadino, and Los Angeles counties, California; and Maricopa and Pima counties, Arizona. (Maricopa and Pima counties include Phoenix and Tucson, respectively). Within each of these large counties, sub-areas will be targeted where demand for the PITC is high.

**Exhibit 9.**  
**Licensed Child Care Programs in Target Counties**

County	Total		Spanish Language	
	Child Care Centers	Family Child Care Homes	Child Care Centers	Family Child Care Homes
Orange	718	1761	352	510
Riverside	311	1880	171	564
San Bernadino	399	1855	271	649
Los Angeles	2230	7823	1315	3677
Maricopa	1034	954	n/a	n/a
Pima	279	85	n/a	n/a

Source: Arizona Child Care Resource and Referral ([arizonachildcare.org](http://arizonachildcare.org)) and 2005 California Child Care Portfolio.

Not all of the providers included in the figures presented in Exhibit 9 serve infants and toddlers, but a large proportion do. In each of these areas a significant proportion of providers predominantly serve children from Spanish-language households. Most day care centers serve a mix of English and Spanish-language children while family day care providers generally serve a single language group. In our sample selection, we will target English and Spanish-language providers, and where necessary we will oversample providers to ensure that approximately half of our sample falls in one of

these two language groups. This will enable us to disaggregate program effects along this important subgroup dimension.

### ***Universe of Children***

The universe of children from which the sample will be drawn encompasses all children enrolled in the above programs who are under the age of 24 months at the time of intake and whose families have a primary language of either English or Spanish.

### ***Sample Design and Selection***

The study uses a two-level, nested sample design and analyzes outcomes both at the level of child care providers and at the level of individual children. Sample sizes are chosen in such a way that impact analyses have sufficient statistical power at the program level and at the child level. (The assumptions underlying the statistical power calculations and the results of these calculations are discussed below). At the center/provider level we will describe PITC's impacts on the child care environment, child-provider interactions, and the level of cognitive stimulation and learning opportunities available to the children. At the child level, we will measure the children's cognitive, language, and socio-emotional development.

The following factors and constraints influence the sample design:

1. The statistical power needed to detect program effects that are meaningful from a policy and cost-benefit perspective.
2. The cost of recruiting, serving, and collecting data from additional child care providers.
3. The need to include sufficient numbers of child care centers and family day care homes to enable separate impact estimates for these two distinct types of child care.
4. The average number of children being cared for in day care centers and by family day care providers.
5. The proportion of children in each of the recruited programs who meet our selection criteria in terms of age and language, and are available for follow-up data collection.

The statistical power analyses presented below address the first three of these considerations and establish sample sizes at the program and child levels that are both sufficiently large and cost effective. However, these calculations represent assumptions regarding the size and composition of the child care settings that may need to be modified after the site recruitment effort gets underway and more details emerge about the nature of recruited providers, the number of children they serve, and our ability to obtain informed consent from parents.

### ***Provider Characteristics***

The recruitment and selection of providers is guided by five considerations. First, as part of the WREL contract, the focus of the study is the Western region, which includes the states of California, Arizona, Nevada, and Utah. The sample will be recruited within these states. Second, to estimate the net impact of PITC it is important that there is not already significant penetration of PITC in the areas from which we recruit. This excludes certain areas in California, which are already being served by PITC through existing arrangements. In such areas providers would either already have been exposed to PITC or would have declined to take advantage of the program previously. Neither of these groups of providers would make good candidates for an experimental test of the program. Third, we want to select programs that are broadly representative of the region and serve children that are similarly representative of the target population for programs like these and for alternative programs such as pre-K and Early Head Start. Fourth, we want the programs to have the potential to benefit from PITC. This means that they need to have sufficient organizational and staff stability so that staff members who participate in PITC have a chance to practice what they learn. It also means that they should not be model programs, whose staff might not need the training and support that PITC provides. Lastly, we plan to recruit the providers in four distinct geographic areas. This enables us to recruit sufficient numbers of providers and conduct implementation and follow-up data collection in a cost-effective manner.

Specifically, we plan to recruit providers in Los Angeles, San Bernadino, Riverside and Orange Counties, and the larger Phoenix and Tucson areas. The four California counties represent areas where PITC has not been widely implemented and have large numbers of both child care centers and family day care providers, many of whom serve predominantly Spanish-speaking children whose early education outcomes tend to lag behind their English-speaking counterparts (Crosnoe, 2005). The Phoenix and Tucson areas share many of the same characteristics and have fast-growing child care industries with limited resources for training and professional development of child care providers. Both Arizona and California also have strong movements to promote early childhood education through universal pre-K programs. The ability of the early care and education system to support the development of children entering pre-K programs is therefore a critical and enduring concern in these regions.

### ***Child Characteristics***

The study targets children who receive child care from a participating provider for at least twenty hours per week and are between 3 and 24 months old at the time of random assignment. This means that the children will be between 15 and 36 months old at first follow-up and between 27 and 48 months old at second follow-up. Younger children will be excluded because we do not expect short-term impacts on them and do not have reliable instruments with which to assess their development at first follow-up. Older children are excluded because they are likely to “graduate” out of infant and toddler care shortly after random assignment and are therefore less likely to experience significant program benefits.

As discussed in more detail below, we do not expect to be able to enroll all age-eligible children in the study. Before conducting random assignment (at the program level), we will seek to obtain parental informed consent for the two rounds of follow-up data collection with participating children. All children for whom we obtain such consent will

be included in the sample. While this creates a somewhat purposive research sample, selecting the children before random assignment minimizes any impact on the validity of our findings. We also will develop an effective communication strategy (see below) to maximize parent cooperation and consent. This will help maximize the size and representativeness of our child sample.

### ***Steps in Sample Selection***

All recruitment, random assignment, and data collection will occur in waves. Approximately 60 programs will be recruited in each of two waves in each of the two states (a total of 240 programs as indicated in Exhibit 4). Steps in each wave will take place as follows:

The research team will establish target numbers of day care centers, family day care providers, and children (including key subgroups such as language subgroups).

Program staff in California and Arizona will obtain lists of licensed child care centers and family child care homes in the designated geographic areas. An initial sample of programs will be drawn. Recruiters will send recruitment flyers and follow up with phone calls. Recruiters will administer a brief screening form and enter answers into a common database. BPA will train recruiters to explain the study to applicants and to administer the screening.

Additionally, in recruitment of family child care homes, the program staff/recruiter may work with provider associations or resource and referral agencies. These agencies may be asked to distribute recruitment flyers and to allow program staff to make presentations at provider meetings.

Researchers will make an initial selection based on results of the screening, and depending on the level of response to our initial contact, follow up with additional outreach efforts if needed. Those selected will be invited to an orientation meeting led jointly by BPA and program staff. Agendas for these meetings will include:

- We will provide an overview of the program and the study.
- We will distribute, explain, and collect the program consent form.
- We will discuss the process for obtaining parent consent: We will distribute parent consent forms; ask providers to distribute and collect these from parents; and give providers guidance in explaining the study to parents. The parent forms include a refusal option as well as a consent option. Parents and providers will be given phone numbers to call with questions.

In order for a child care center to be included in the study, two infant/toddler classrooms (or one classroom if the center has only one infant/toddler classroom) must participate. In addition to the director's signed consent, consent from at least two caregivers/teachers per classroom and parents of three to five children per classroom will be needed. In order for a family child care home to be included in the study, consent from the director/owner as well as from a minimum of two parents of infant/toddlers will be needed. Research staff will follow up with providers about the parental consent process for approximately three weeks. After three weeks the provider will either be dropped from the sample or study participation begins. It is important to limit the time that expires

between parental consent and random assignment to minimize child turnover and sample attrition.

## **2. Statistical Power of the Sample**

The purpose of our sample design is to produce impact estimates that have sufficient statistical precision so that impacts that are practically meaningful will also be statistically significant. Thus, for example, if providing PITC training and support to a child care center would cost \$300 per child per year, a meaningful program impact would be one that is comparable or larger than those found in other evaluations of similarly priced early childhood interventions. Conversely, a significantly smaller impact would arguably not be policy relevant and our study would not need to have sufficient statistical power to detect it.

Expressed in terms of effect sizes (the impact of an intervention divided by the standard deviation of the outcome), early childhood educational interventions are usually found to have relatively larger impacts than interventions that target children later in life (Borman et al., 2003). Exhibit 4 shows the expected effect sizes for our study design, which assumes samples of 90 child care centers (with 8 children each, on average) and 150 family child care providers (with 3 children each, on average<sup>1</sup>). With these sample sizes, which were chosen during initial power calculations and reflect cost and logistical constraints, the study would be able to detect effect sizes of 0.28 to 0.32 for child care centers and 0.24 to 0.28 for family day care providers. For the full sample, the child-level MDES would be between 0.18 and 0.22. (Details underlying these calculations are discussed below).

At the provider level our study has less statistical power. Provider-level MDES are estimated at 0.48 for day care centers, 0.37 for family day care providers, and 0.33 for the full sample. This means that changes in the child care environment and staff knowledge need to be quite substantial to be statistically significant. However, for a program like PITC to have effects in the order of 0.25 standard deviations on distal child outcomes, one might expect to need effects in the range of 0.35-0.45 on proximal provider-level and environmental outcome variables.

---

<sup>1</sup> There is some concern that it may be difficult to recruit sufficient numbers of family day care providers serving as many as three infants or toddlers. In California, the maximum number of these children that can be cared for by a family child care provider is four. If it turns out that we need to include a significant number of providers with two infants or toddlers, we will increase our sample size of family day care providers accordingly. Doing so would increase the power of our study, because we would increase the number of clusters while lowering the average number of children in each cluster. However, it would also increase the cost of conducting data collection, which might require adjustments in the scope of data collection and analysis.

**Exhibit 10.**  
**Statistical Power Calculations**

<b>MDES for child outcomes (cluster random assignment)</b>					
	J	n	N*	MDES 1	MDES 2
Full sample	240	4.9	936	0.20	0.24
Child care centers	90	8	576	0.25	0.33
Family child care homes	150	3	360	0.29	0.33

  

<b>MDES for provider outcomes (simple random assignment)</b>			
	n	N*	MDES
Full sample	240	216	0.37
Child care centers	90	81	0.56
Family child care homes	150	135	0.48

Source: Calculations using Optimal Design Software (Raudenbush and Liu, 2000)

Notes: J = Number of clusters, n = Number of children per provider, N = number of follow-up data points. MDES 1 = MDES for ICC of 0.1, MDES 2 = MDES for ICC of 0.2, MDES = Minimum Detectable Effect Size.

\*We estimate a 20 percent attrition rate for children and a 10 percent attrition rate for providers. Note that these two attrition rates are estimated separately. We will follow up with individual children even if we cannot conduct observations at the provider level.

Given the existing effect sizes found in evaluations of interventions for young children, it might have been tempting to design our study with an MDES of 0.4 at the child level. However, the existing research typically reports on interventions that are more costly and more multi-faceted than the PITC program, and that more directly intervene in the lives of young children and their families. This would suggest that the PITC, which is a more modest intervention, might be considered successful and cost effective if it produces impacts that are smaller than those commonly found in evaluations of early childhood interventions. This in turn would suggest that it is appropriate to strive for MDES smaller than 0.4. An added advantage of doing so is that it creates a “margin of error” so that we are unlikely to experience insufficient power even if there were to be more variation in outcome measures or more sample attrition than expected. Lastly, MDES larger than those currently expected at the provider level would make it increasingly unlikely that we would be able to reliably detect program-level impacts.

#### **Other considerations for the statistical power estimates**

In cluster random assignment evaluation designs like this one, a critical factor in the estimation of statistical power is the extent to which individual observations are independent of one another within the units of random assignment (in this case the child care programs). A high degree of clustering of outcomes within these units reduces statistical power and a high degree of independence increases it. This ratio of between-cluster variance and within-cluster variance is captured by the intra-class correlation coefficient.

There is little relevant research on the intraclass correlation of child outcomes within child care settings. In designing their National Head Start Evaluation, Puma et al. (2005) found a very large intra-class correlation of 0.51 for child outcomes within child care settings using data from the Head Start Family Experiences Survey (FACES). However, that data is from a national survey and reflects significant geographic and socio-economical clustering, which is less likely to be a major factor within our more geographically contained sample of providers. Puma et al. also argue that a high degree of intraclass correlation in outcome *levels* does not mean that the degree to which a program *improves* test scores will associate so strongly between children served by the same providers. They then go on to use an intraclass correlation of 0.20 or 0.30 for their impact analyses, but admit that the lack of empirical evidence makes this choice rather arbitrary. Another source of early childhood education ICC estimates is a recent paper by Schochet (2005), who finds ICCs of approximately 0.2 in preschool settings.

As Bloom et al. (1999) showed, the detrimental effects of clustering on statistical power can be mitigated by including group- or individual-level baseline control variables in the impact analysis. These variables, measured before random assignment, reduce the impact of clustering by removing much of the random variation in background characteristics between clusters. For this to work well it is important that the covariate being used is a strong predictor of the outcome variables being studied. In schools, it is often possible to use either a pre-test for students who receive an intervention or use test scores for a previous cohort of students. It is unlikely that we will have such data at the provider level in our study. As a result we expect the benefits of using covariates to reduce the impact of clustering to be smaller than they would be in a school setting. (We describe the impact analysis methods in more detail in Section 5.3 of this research plan.) For the purpose of calculating minimum detectable effect sizes we are using two different assumed levels of the intra-class correlation, namely 0.1 and 0.2. With limited ability to control for child background characteristics, we conservatively assume that we will be able to reduce the effects of clustering to at least match an ICC of 0.2. If our sample of providers is reasonably homogeneous or if we are better able to predict cross-provider variation in child outcomes, the scenario with an ICC of 0.1 would be more appropriate.

### **3. Maximizing Response Rates**

Estimated retention rates for the study are eighty percent for children and ninety percent for providers. These rates represent the average expected response rate after the initial enrollment in the study. Research partners for this study have a record of success in producing high response rates and high quality, reliable data on children and families. The research team is committed to careful training and oversight of all field data collection staff, and will maintain high inter-observer reliabilities through ongoing cross-site checking.

Among strategies that will be used to ensure high quality data and high response rates are the following: 1) We will provide training, oversight, and data collection manuals for all data collection staff; 2) We will compensate respondents and maintain contact with them in between rounds of data collection; 3) We will maintain up-to-date contact information for all study participants, including multiple forms of contact; 4) We will track responses in an integrated project data system and conduct extensive follow-up with non-respondents; 5) We will consult with PITC staff regarding lower than expected

response rates, crossovers, or other difficulties. Senior research and project staff will meet monthly to discuss any problems and revise protocols or intervene with sites as needed.

Specific methods to be used to maintain contact with each group of respondents are described below. These methods have been used successfully by members of the research team in previous studies such as The Study of the New Chance Demonstration Program for Young Mothers in Poverty and the Milwaukee Family Study/New Hope Study.

- Participating parents will receive a postcard every 4-6 months thanking them for their participation and requesting that the Study team be notified of any changes of address or other contact information.
- Participating children will be mailed birthday cards.
- Participating child care programs will be contacted by telephone at the mid-point between the two rounds of program data collection (winter 2008). Any major changes in program staffing or location will be noted at this time.
- Program directors will have several early contacts with the study team-- including in-person, mail, and telephone contacts-- during the recruitment/enrollment period and prior to baseline data collection (See Steps in Sample Selection above.) We will include a letter to program directors confirming enrollment and notifying them of upcoming baseline data collection activities.
- Parents and child care providers will receive a letter approximately one month in advance of follow-up data collection activities, reminding them that they will be contacted soon regarding upcoming questionnaires and observations.

### ***Program/Caregiver Observations and Questionnaires***

Program and caregiver observations will be conducted by trained field researchers, working as a team under the supervision of a field coordinator at Berkeley Policy Associates. For each of the two rounds of program observations/questionnaires, a researcher will spend a full day of on-site data collection with each family child care home and up to two days of on-site data collection with each center (depending on the size of the center and the number of infant/toddler classrooms.) Questionnaires will be sent to programs two weeks in advance of the visit, with the request that caregivers complete these and submit to the researchers at the time of the visit. Most of the on-site time will be devoted to completing observation protocols, with one-half hour remaining for the researcher to ask questions that have not been completed in advance of the visit.

Program-level data collection staff will be hired and managed by Berkeley Policy Associates, with training and oversight coordinated by the University of Texas. All observers will participate in a one-week training in January 2007, and will receive a training manual and training videotapes. Inter-rater reliabilities will be established at a minimum of 80% agreement for all observational measures prior to data collection; inter-

rater reliability will be re-checked during later waves by having the field coordinator accompany observers in each wave to re-establish reliability.

### ***Child Outcomes Data Collection***

Child data collection staff will be hired and managed by SRM Boulder, with training and oversight coordinated by the University of Texas. All child data collection staff will have some previous undergraduate or graduate-level training in child development, and will also participate in a three-day training session to establish expertise and inter-rater reliability in the measures used for the study. Inter-rater reliability will be established to reach 80% agreement and will be re-established in each state at mid-course through each round of data collection.

Researchers will meet with children in their child care settings at a time when the parent can be present, if possible; alternatively, they will arrange to meet children with parents in their homes. Researchers will carry a small portable room divider and toys in order to create a consistent environment, without distractions, in which they can privately assess each child. Parents will be mailed questionnaires in advance of the child visits and will be asked to complete them by the time of the visit if possible; if not, the researcher may assist the parent in completing the questionnaire during the visit.

The research team will maintain ongoing contact with participating families, facilitated by gathering several alternative phone numbers as well as email and mail contacts at the start of the study. Families will receive reminder letters beginning several months in advance of each data collection round, at which time researchers will contact them by phone and email for scheduling.

### ***Payments to Respondents***

Payments to study respondents will be used to offset respondent burden, to maintain a positive relationship with respondents, and to maximize response rates.

Study participants will be paid as follows:

- **Programs:** For child care centers, we will provide a \$15 gift card per classroom (a maximum of two classrooms per center will be included) to each program that returns the completed packet of caregiver and parent informed consent forms within two weeks. A completed packet will include a minimum of two caregiver forms per classroom and three to six (depending on classroom size) parent forms per classroom. For family child care homes, we will provide a \$15 gift card for each home that submits completed informed consent forms from all parents of enrolled children under the age of twenty-four months. Again, all forms are counted, including those that indicate refusal to participate. **All parent forms are counted including those that indicate refusal to participate.**
- **Caregivers/Teachers:** Each individual caregiver will receive \$25 merchandise gift cards for each questionnaire completed (one in 2007 and another in 2008).

- **Families:** Families will receive \$10 merchandise gift cards after completion of a consent form with a baseline questionnaire and \$50 after each in-person research session (one in 2008 and another in 2009).

Similar payments have been used in comparable studies conducted by members of the research team. In the Milwaukee Family Study (also called the New Hope Study), conducted by the Manpower Demonstration Research Corporation and the University of Texas with SRM Boulder, parents were paid \$50 for participation in each parent-child interview session, and children over the age of six were given gift coupons worth \$15-\$20.

Caregivers in the program group who complete all requirements of the PITC training will receive professional growth compensation in the form of either academic units, or \$350 in the form of cash or resource materials. This compensation is part of the PITC intervention and are not specifically related to participation in the study.

### ***Generalizability and External Validity Checks***

As part of the impact analysis we will compare characteristics of our sample programs and children to those of the universe of programs and children in the region, to the extent possible. Data on children's demographic and household characteristics by county are available through the 2000 Census. While child care quality data are not available at the county or state level, child care supply data published by resource and referral agencies include limited characteristics of programs and providers such as language and hours. Also for an assessment of generalizability, we will use comparative data on quality and characteristics of child care programs and caregivers as reported in major research studies, some of which include sites in California and Arizona (Whitebook et al, 1994; Cost, Quality, and Outcomes Study Team, 1995; Galinsky, et al.,1994; NICHD Early Child Care Research Network, 2005).

## **4. Pretesting**

The study makes use of instruments (or selections from instruments) that have already been extensively tested and fielded in large studies such as the Early Childhood Longitudinal Study and the NICHD Study of Early Child Care and Youth Development. We will conduct limited pretesting of instruments designed specifically for this study, to ensure that the respondent burden does not exceed our estimates. Fewer than ten respondents will be included in the pretest.

## **5. Contact Information**

Key individuals contacted on statistical aspects of the design include:

Hans Bos, CEO and Co-Principal Investigator, Berkeley Policy Associates, 510-465-7884 x 217, [Hans@bpacal.com](mailto:Hans@bpacal.com)

Neal Finkelstein, Senior Research Scientist, WestEd, 415-565-3000,  
[nfinkel@wested.org](mailto:nfinkel@wested.org)

Tom Hanson, Senior Research Associate, WestEd, 415-565-3000, [thanson@wested.org](mailto:thanson@wested.org)

Aletha Huston, Professor and Co-Principal Investigator, University of Texas, 512-471-0753, [achuston@mail.texas.edu](mailto:achuston@mail.texas.edu)

For more information about the conduct of the study, contact:

Phyllis Weinstock, Project Director, Berkeley Policy Associates, 510-465-7884 x205,  
[Phyllis@bpacal.com](mailto:Phyllis@bpacal.com)

## References for Part B

Bloom, H. S., Bos, J. M., & Lee, S. (1999). Using cluster random assignment to measure program impacts. *Evaluation Review*, 23(4), 445–469.

Borman, G.D., Hewes, G.M., Overman, L.T., & Brown, S. (2003). Comprehensive School Reform and Achievement: A Meta-Analysis. *Review of Educational Research*, 73, 125-230.

Cost, Quality & Child Outcomes Study Team (1995) *Cost, Quality and Child Outcomes in Child Care Centers, Public Report*, Second Edition. Denver: Economics Department, University of Colorado at Denver.

Crosnoe, R. (2005) "Double Disadvantage of Signs of Resilience: The Elementary School Contexts of Children from Mexican Immigrant Families." *American Educational Research Journal* 42:269-303.

Galinsky, E., C. Howes, S. Kontos, & M. Shinn. 1994. *The study of family child care and relative care: Highlights of findings*. New York: Families and Work Institute.

NICHD Early Child Care Research Network (2005). *Child Care and Child Development Results from the NICHD Study of Early Child Care and Youth Development*. New York: Guilford Press.

Puma, M., S. Bell, R. Cook, C. Heid, & M. Lopez (2005). "Head Start Impact Study: First Year Findings." Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families.

Raudenbush, S.W., & Liu, X.F. (2000). "Statistical power and optimal design for multisite randomized trials", *Psychological Methods* 5(2): 199-213.

Schochet, P. (2005). Statistical power for random assignment evaluations of education programs. Document No. PR05-36. Princeton, NJ: Mathematica Policy Research.