



AGENDA ITEM: 9
DATE OF MEETING: April 24, 2014
ACTION: X
INFORMATION: _____

TRANSITIONAL KINDERGARTEN EVALUATION

SUMMARY

Per Commission request, staff is providing information regarding the feasibility of data collection for Asian languages and wraparound services as supplements to a Transitional Kindergarten (TK) evaluation. This information was requested at the January 23, 2014, Commission meeting and was prepared by the American Institutes for Research (AIR). The attached AIR Research Memorandum discusses the feasibility of various options.

Staff requests Commission approval of \$1 million to complement the \$500,000 approved at the January 23, 2014, Commission meeting for a total of \$1.5 million to fund a co-partnered evaluation, with the Heising-Simons and Packard Foundations, of California's implementation of TK during the three-year period, January 2014 through December 2016. The attached AIR Research Proposal describes in detail the research questions, methods, data collection, work plan, timeline, and total budget of \$7.4 million.

BACKGROUND

First 5 California was invited by the Heising-Simons Foundation and Packard Foundation to partner in the funding of an evaluation of California's implementation of TK. AIR was awarded an evaluation contract by the lead funding partner, Heising-Simons. The cost of the study is approximately \$7.4 million. Heising-Simons will be the major funder contributing up to \$4.4 million, with Packard and First 5 California participating as secondary funders at \$1.5 million each.

ATTACHMENTS

- AIR Research Memorandum: ***Transitional Kindergarten Outcomes Study Proposal Addendum: Feasibility of Including Assessments in Asian Languages and Information About Wraparound Services.*** American Institutes for Research, April 24, 2014.
- AIR Research Proposal: ***Quality and Outcomes Study of California's Transitional Kindergarten Program, Technical Proposal.*** American Institutes for Research, September 27, 2013.

Transitional Kindergarten Outcomes Study Proposal Addendum:

Feasibility of Including Assessments in Asian Languages and Information About Wraparound Services

April 24, 2014

Submitted to: First 5 California
California Children and Families Commission
2389 Gateway Oaks Drive, Suite 260
Sacramento, CA 95833

Submitted by: American Institutes for Research
1000 Thomas Jefferson Street NW
Washington, DC 20007
Nilva da Silva
202-403-5086
E-mail: ndasilva@air.org

The purpose of this memo is to provide First 5 California with American Institutes for Research's (AIR's) assessment of the feasibility of addressing, within the Transitional Kindergarten Quality and Outcomes Study¹, two concerns of the California Children and Families Commission (the Commission). This document presents each of these concerns, including the estimated cost of proposed optional study components.

Concern 1: Asian/Pacific Islander students who do not speak English proficiently are not included in direct child assessments in the study.

Asian/Pacific Islander students are included in the Transitional Kindergarten (TK) study in two ways. First, their experiences in TK classrooms are measured according to the Classroom Assessment Scoring System (CLASS) and Emerging Academic Snapshot tool. The CLASS tool has been shown to be valid and reliable, and predictive of child outcomes, even in classrooms with high proportions of dual-language learners (Fuligni & Howes, 2011; Downer et al., 2011). In addition, all age-eligible Asian/Pacific Islander students who can be tested in English (i.e., who pass the oral language screener given at the beginning of the direct child assessment session) will be invited to participate. Also, parent letters and consent forms will be translated into Asian languages according to the needs of and population in each participating district.

How many students will be left out if assessments in Asian languages are not included?

Students who are not sufficiently proficient in English to be assessed in English are not included in child assessments in the current study design. However, only a very small proportion of the overall student population (approximately 3 percent statewide) would be ineligible to participate because they are limited in their English proficiency (and speak a home language other than Spanish). Based on California Department of Education (CDE) data on kindergartners in 2011–12, we estimate that two thirds of the age-eligible Asian/Pacific Islander student population could be tested in English and therefore included in the study (see Table 1).

What would it take to test students in one or more Asian languages?

There are currently no valid direct child assessment measures in literacy and math available in languages other than English and Spanish that would be appropriate for use in the TK study. To develop such assessments would be a major assessment development effort, very expensive, and essentially infeasible in the time period available (as child assessments will begin in September 2014). Translating and validating the new measures could not be done in time for data collection on students in Cohort 1 (fall and spring of the 2014–15 kindergarten year), and the validation work could interfere with the Cohort 1 data collection by overburdening districts already selected for the study. The number of students required to validate a single translation also would make validating multiple translations cost prohibitive and time prohibitive.

¹ The Heising-Simons Foundation is currently supporting AIR in conducting a study of outcomes for students who attend Transitional Kindergarten (TK), and to examine the relationship between characteristics and the quality of TK programs and these student outcomes in kindergarten. Please see the larger study proposal included with this memo for the April 24, 2014, Commission meeting.

Table 1. Proportion of Asian/Pacific Islander Students in California Estimated to Be Ineligible for Direct Assessments as Part of the TK Study Because of Limited Proficiency in English

	Total Number of Kindergarten Students in California, 2011–12	Estimated Number of Students Age-Eligible for the TK Study (born Oct 1–Feb 2) in California	Estimated Number of Students Sampled for the Study
Students identified as ethnically Asian/Pacific Islander			
Total	48,645	16,215	269
Designated as English learners (ELs)	22,656	7,552	125
Students scoring below <i>Early Advanced/Advanced</i> on California English Language Development Test (CELDT), by language group			
Vietnamese	5,076	1,692	28
Cantonese	2,570	857	14
Mandarin (Putonghua)	2,511	837	14
Korean	2,121	707	12
Tagalog or Pilipino	2,108	703	12
Hmong	1,340	447	7
Khmer	547	182	3
Summary			
Total Asian/Pacific Islander students ineligible for assessment in English	16,273	5,424	90
Percentage of total Asian/Pacific Islander students	33%	33%	33%
Percentage of total kindergarten population	3%	3%	3%

Source: California Department of Education, 2012–13; California Basic Educational Data System, 2012–13 CELDT Research File

Furthermore, even if measures could be developed, our impact analyses would not be able to show results disaggregated by language subgroup; we would need at least 2,000 students in each subgroup to have sufficient statistical power to analyze and present results for that particular group. Instead, results would have to be included with those of other ethnic subgroups. In other words, Asian/Pacific Islander children would be included in the study and in statements about the impact of TK overall, but we could not describe the impact of TK on Asian/Pacific Islander students as a particular subgroup.

Alternatives

There are two options for including Asian/Pacific Islander students who are not proficient in English in the study, which are more feasible.

A. Focus on measures of social-emotional development. Although translating direct assessments of children’s early literacy and mathematics development is infeasible for this study, we could address the question of how Asian/Pacific Islander TK students’ social-emotional skills in kindergarten compare with those of their peers, using translated or teacher-reported measures. Using such measures would address the following research question:

Does TK participation improve children’s social-emotional skills and executive functioning in kindergarten?

First, the planned measure of children’s executive functioning, the Head-Toes-Knees-Shoulders (HTKS) task, could be translated into Asian languages and fairly easily pilot-tested and administered. The HTKS assessment is a set of instructions given verbally to children, and children’s responses (actions) are recorded to score the assessment. These instructions could be translated into one or more Asian languages. In fact, this measure has already been translated into Chinese languages for the recent San Francisco Preschool for All evaluation. In contrast, developing math and literacy assessments in an Asian language is not just a matter of translation, particularly as phonological awareness and other language and literacy skills are different in different languages and cultures; therefore, content would have to be modified and tested in addition to having the language translated. In addition, a second planned measure of social-emotional development involves teacher reports of students’ social behaviors. No translation of this measure would be necessary, and results for Asian/Pacific Islander students could be included in the report. Note that because the sample size of Asian/Pacific Islander children would be so small, presenting results regarding executive functioning skills for Asian/Pacific Islanders alone would not be possible; results would only be reported for the entire group of kindergartners, but with Asian/Pacific Islander students included.

The cost of translating the HTKS task into Mandarin, Cantonese, Tagalog, Korean, and Vietnamese (the five most common Asian languages according to CDE data); piloting the new measures to ensure the instructions are appropriately translated to be valid and reliable while making sense to the target children; recruiting additional children for the sample so that the group of children being assessed on all measures is still at least 3,000; and hiring additional assessors to travel around the state and administer the five-minute assessment is estimated to be approximately **\$85,000–\$100,000**.

B. Focus on English Language Development (California English Language Development Test [CELDT] data). As an alternative to AIR-administered direct assessments to measure the progress of these students, we could obtain and analyze CELDT² scores for all Asian/Pacific Islander English learners (ELs) in the state. This approach would go beyond examining outcomes for students in the current sample, but it would provide even more robust information about the English proficiency levels for this group of students and the impact of TK on these proficiency

² We anticipate that CELDT data from 2014–15 will be available in spring 2015.

levels. We could complete this analysis for other EL groups in the state as well (such as Spanish speakers) for comparison. This analysis would answer this research question:

Does TK participation improve kindergarten readiness in the domain of English proficiency, and does this impact vary by language group (e.g., specific Asian languages, Asian languages compared with Spanish)?

This alternative would rely on obtaining student-level data from CDE. CDE has agreed this analysis is possible, pending CDE's completion of its data request process.

The cost of working with CDE to obtain these data, cleaning the data files, conducting the analysis, and adding the findings to the final report is estimated to be approximately **\$55,000–\$70,000**.

Concern 2: Students may have received ancillary, wraparound services that will not be captured in the study.

As a part of the study already under way, AIR is planning to use a brief survey (attached to the study consent form) to ask parents a few questions about their child's prior experience in preschool or child care, in addition to two or three questions gathering basic demographic information. Any additional questions could jeopardize our participation rate; parents are likely to ignore a long list of questions and decide not to return their form. Therefore, we do not recommend adding a long set of questions regarding wraparound services to the parent consent form. Furthermore, this consent form has already been developed and is being distributed to schools, so modifications could be made only to the consent form for the second cohort of students in the study (beginning kindergarten in 2015–16).

However, to collect additional information regarding wraparound services to address the Commission's concern, there is another option. We could gather information from district administrators through the existing survey about the wraparound services provided by the district. This effort would not produce specific information about which wraparound services individual students or families received, or what services were available or received from organizations outside the district, but it would provide some useful contextual information about district-provided services. This analysis could address the following research question:

How do student outcomes vary by district characteristics, such as the types of wraparound services the district makes available to students?

Similar to the parent consent form, the district survey already has been developed and is being administered, so any modifications could be made only for the second cohort of the study (spring 2015). There would be no additional cost to add one or two questions about wraparound services to the district survey.

Summary

Table 2 summarizes how each concern is or can be addressed within the study.

Table 2. First 5 California Concerns and Recommended Options for TK Evaluation

Concern	How Addressed	Research Questions Addressed	Estimated Additional Cost
<p>Concern 1: Including Asian/Pacific Islander students more fully in the study</p>	<p>Current Design: 1. Classroom observations 2. Child assessments in English</p>	<ul style="list-style-type: none"> • How do TK programs, serving children of all language backgrounds, fare on measures of classroom quality? • Does TK participation improve kindergarten readiness in the domains of early literacy/language, and mathematics skills, for students who are proficient in English or Spanish, and to what extent are the impacts of TK sustained through the end of kindergarten for these students? • How do student outcomes vary with TK classroom quality? 	
	<p>Alternative A: Collect data on executive functioning and social-emotional development: translate the Head, Toes, Knees, Shoulders (HTKS) task and include Asian/Pacific Islander children in teacher ratings of social skills.</p>	<ul style="list-style-type: none"> • Does TK participation improve children’s social-emotional skills and executive functioning in kindergarten? <p>(Note that we would <i>not</i> be able to make statements about the social-emotional or executive functioning skills of Asian/Pacific islander students alone.)</p>	<p>\$85,000-100,000</p>
	<p>Alternative B: Analyze state-level California English Language Development Test (CELDT) data to compare English proficiency scores for Asian/Pacific Islander students who were eligible for TK, compared to those not eligible.</p>	<ul style="list-style-type: none"> • Does TK participation improve kindergarten readiness in the domain of English proficiency, and does this impact vary by language group (e.g., specific Asian languages, Asian languages compared with Spanish)? 	<p>\$55,000-\$70,000</p>
<p>Concern 2: Wraparound Services</p>	<p>Current Design: Not addressed</p>		
	<p>Alternative: Add 1–2 questions regarding district-offered services to the existing district administrator survey (cohort 2, spring 2015, <i>only</i>)</p>	<ul style="list-style-type: none"> • How do student outcomes vary by district characteristics, such as the types of wraparound services the district makes available? 	<p>None</p>

References

- Downer, J. T., Lopez, M. L., Grimm, K., Hamagami, A., Pianta, R. C., & Howes, C. (2011). Observations of teacher-child interactions in classrooms serving Latinos and dual language learners: Applicability of the Classroom Assessment Scoring System in diverse settings. *Early Childhood Research Quarterly, 27*(1), 21–32.
- Fuligni, A. S., & Howes, C. (2011). Experiences of low-income dual language learning preschoolers in diverse early learning settings. In C. Howes, J. T. Downer, & R. C. Pianta (Eds.), *Dual language learners in the early childhood classroom*. Baltimore: Brookes.



AMERICAN INSTITUTES FOR RESEARCH®

Quality and Outcomes Study of California's Transitional Kindergarten Program

TECHNICAL PROPOSAL

September 27, 2013

Submitted to: Deanna Gomby, Ph.D., M.S.
Vice President, Education
Heising-Simons Foundation
300 Second Street
Los Altos, CA 94022
E-mail: deanna@heisingsimons.org

Bernadette Sangalang, Ph.D., M.S.W.
Children, Families, and Communities (CFC) Program
The David and Lucile Packard Foundation
343 Second Street
Los Altos, CA 94022
E-mail: bsangalang@packard.org

Camille Maben
First 5 California
2389 Gateway Oaks Drive, Suite 260
Sacramento, CA 95833
E-mail: cmaben@ccfc.ca.gov

Submitted by: American Institutes for Research
1000 Thomas Jefferson Street, NW
Washington, DC 20007
Nilva da Silva
202-403-5086
E-mail: ndasilva@air.org

Tax Identification Number: 25-0965219

This proposal includes proprietary and business confidential data and shall not be disclosed outside the Client and shall not be duplicated, used, or disclosed—in whole or in part—for any purpose other than to evaluate this proposal. However, if an agreement is awarded to this offeror as a result of—or in connection with—the submission of these data, the Client shall have the right to duplicate, use, or disclose the data to the extent provided in the resulting agreement. This restriction does not limit the Client's right to use the information contained in these data if they are obtained from another source without restriction. Notice of Trademark: "American Institutes for Research" and "AIR" are registered trademarks. All other brand, product, or company names are trademarks or registered trademarks of their respective owners.

American Institutes for Research

1000 Thomas Jefferson Street NW, Washington, DC 20007-3835 | 202.403.5000 | TTY 877.334.3499 | www.air.org

Quality and Outcomes Study of California's Transitional Kindergarten Program

REVISED September 2013



AMERICAN INSTITUTES FOR RESEARCH®
1000 Thomas Jefferson Street NW
Washington, DC 20007-3835
202-403-5000 • TTY 877-334-3499
www.air.org

Contents

Introduction.....	48
General Approach.....	49
The TK Logic Model	49
Research Questions	51
Proposed Methods and Analysis.....	52
Identification Strategy: Regression Discontinuity	52
Sample, Data Collection, and Measures	53
Analytic Approach	59
Workplan and Timeline	63
Task 1: Planning, Preparation, and Training	63
Task 2: Sampling and Collecting Student Record Data.....	65
Task 3: Recruitment.....	66
Task 4: TK Classroom Observations	67
Task 5: TK Teacher Survey	68
Task 6: District Survey and Follow-Up Interviews	69
Task 7: Kindergarten Student Assessments	69
Task 8: Kindergarten Teacher Survey	70
Task 9: Data Analysis	71
Task 10: Reporting.....	72
Task 11: Project Management	73
Staff and Organizational Experience	75
AIR’s Institutional Review Board	75
Budget	75



AMERICAN INSTITUTES FOR RESEARCH®

September 27, 2013

Deanna Gomby, Ph.D., M.S.
Heising-Simons Foundation
300 Second Street, Suite 240
Los Altos, CA 94022

Bernadette Sangalang, Ph.D., M.S.W.
Children, Families, and Communities (CFC) Program
The David and Lucile Packard Foundation
343 Second Street
Los Altos, CA 94022

Camille Maben
First 5 California
2389 Gateway Oaks Drive, Suite 260
Sacramento, CA 95833

RE: *Quality and Outcomes Study of California's Transitional Kindergarten Program*

Dear Dr. Gomby, Dr. Sangalang, and Ms. Maben:

American Institutes for Research (AIR) is pleased to submit its revised proposal, *Quality and Outcomes Study of California's Transitional Kindergarten Program*, for your consideration.

In response to your feedback, we have made a number of revisions, including:

- Adding spring assessments and kindergarten teacher surveys (and associated training, management, data analysis and reporting time) for Cohort 2
- Proposing a set of direct assessments that address the constructs of interest and that can be completed within an assessment session of approximately 45 minutes
- Giving additional consideration to issues for dual language learners, including ensuring that all measures are available in Spanish and are culturally appropriate; we also will add an additional TAG member with expertise in this area
- Adding a set of regression discontinuity graphic representations to help the reader visualize the anticipated analysis
- Adding more information on our subcontractor, Survey Research Management (SRM), regarding their experience with this kind of work and their budget

We are e-mailing one electronic version of the proposal, as requested. Please direct contractual questions about this proposal to Nilva da Silva, Contracts Officer, at 202-403-5086 or ndasilva@air.org. For technical questions, please contact Heather Quick at 650-843-8130 or hquick@air.org.

Thank you for the opportunity to submit this proposal. We look forward to working with you on the next phase of this important study.

Sincerely,

Hans Bos
Vice President

Introduction

In 2010, Governor Arnold Schwarzenegger signed the Kindergarten Readiness Act (SB 1381) into law. The law changed the kindergarten age cutoff from December 2 to September 1, phasing in the new age requirement by moving the cutoff date back one month per year for three years, beginning in fall 2012. SB 1381 also established a new grade level—transitional kindergarten (TK)—which is the first year of a two-year kindergarten experience for students born between September 2 and December 2. When fully implemented, TK is intended to provide an additional year of early education to 120,000 children each year, with the goal of promoting their school readiness and reducing the achievement gap.

This major shift in education policy in the state has grabbed the attention of stakeholders at all levels of the educational system—from early childhood education advocates who view this as an important step toward broader access to early education experiences to community colleges now working on supplementing their teacher education programs with coursework for TK teachers. There is significant interest in understanding not only what this new program looks like in practice but how it supports students' learning, development, and readiness for kindergarten.

In 2012, the Heising-Simons Foundation and The David and Lucile Packard Foundation led the effort to examine the rollout of TK by funding American Institutes for Research (AIR) to conduct an implementation study of the new program and to develop a plan for a study to assess the quality of TK and its impacts on student learning. AIR, a not-for-profit corporation with more than 60 years of experience in education-related research and development projects, is pleased to submit this proposal to the Heising-Simons Foundation, The David and Lucile Packard Foundation, and First 5 California to conduct the second phase of the TK study: Quality and Outcomes Study of California's Transitional Kindergarten Program.

We propose a strong team of researchers and analysts led by Dr. Heather Quick, Principal Investigator, and Karen Manship, Project Director, who, together, successfully led the implementation study. Dr. Hans Bos and Deborah Parrish will serve as senior advisors, offering content and methodological guidance over the course of the study. Supplementing the implementation study team will be experts in the regression discontinuity (RD) method and the direct assessment of young children, including our subcontractor, Survey Research Management (SRM), which has many years of experience successfully conducting large-scale data collections such as that proposed here.

Building on the implementation study findings, the quality and outcomes study will focus on three main goals. First, we will measure the quality of TK classrooms, using direct observations of instruction in combination with teacher reports of their approach to TK and classroom practices. Second, using an RD design, we will assess the impact of TK on student readiness for kindergarten by using direct student assessments of language, literacy, mathematics, and social-emotional development and by comparing outcomes for children who attended TK with those children who missed the cutoff and were too young for TK. Finally, we will examine the relationship between the TK program quality and outcomes for TK students.

AIR brings a unique combination of strengths and relevant experience to this proposed study. First, AIR has a strong group of early care and education researchers, many of whom have extensive experience in evaluating and analyzing prekindergarten programs and other school readiness programs in California as well as in serving as research and policy advisors to statewide entities, such as the State Advisory Council on Early Learning and Care. AIR has a long history in California and in the evaluation of educational programming for young children in the state. We understand the California policy context, the vision behind the TK program, and implementation issues and challenges that teachers and administrators confront each day.

Second, in addition to the knowledge base we have built around TK through the implementation study, we have also developed relationships with many districts and statewide organizations focused on supporting TK, which will facilitate our outreach to districts and the successful recruitment of districts, schools, teachers, and families into the study. Our experience studying TK as well as conducting other state and local evaluations will help us quickly understand the landscape of the staff and administrators in the public school settings and how to engage them in the study. In each of these evaluations, we have demonstrated our success in engaging teachers, administrators, and parents to ensure high rates of participation. We employ a continuous feedback loop wherever possible, providing information on the status of the study to key stakeholders and participants as well as sharing findings through the course of the study.

Third, AIR brings evaluation design expertise and utilizes state-of-the-art outcome evaluation strategies that are tailored to the questions at hand, resulting in information that is valid, reliable, and relevant to programs, policymakers, and other stakeholders. AIR has specific expertise in impact evaluations, both randomized controlled trials and studies employing an RD design, and offers a skilled team guided by methodologists and technical experts to ensure the study's success. In addition to design strengths, AIR has expertise in assessing child outcomes using a variety of direct child assessments and teacher observations and is very familiar with elementary school data sources and standardized test databases; we have often used methods and data sources to assess program impact. We also have a breadth of experience measuring classroom quality, having conducted observations using the Classroom Assessment Scoring System (CLASS) in many contexts and for many years, even before its official release, as well as other program quality measurement tools.

General Approach

AIR's approach to the TK quality and outcomes study emphasizes expertise, experience, and engagement. We will draw on in-house and external experts to ensure that content, design, and methodological decisions are well considered. This will include calling on the technical advisory group (TAG) members identified for the implementation study as well as additional experts as needed. Specifically, we will augment the TAG with at least one additional expert in issues related to English learners. We will build on the findings from the implementation study, which identified a number of quality dimensions that differentiated TK from kindergarten classrooms (e.g., a focus on social-emotional development versus an emphasis on academic content areas). We will use lessons learned from the implementation study to guide the development of data collection tools and also the procedures for carrying out the data collection activities. Finally, we will use a participatory evaluation approach, engaging the study stakeholders by holding up-front conversations with districts about the study processes and by sharing results with stakeholders—including district staff, principals, teachers, advocates, and policymakers—through research briefs and district-specific reports of findings. Sharing emerging results throughout the course of the study will inform quality improvement efforts among districts and schools, give stakeholders the opportunity to help us interpret results, and facilitate buy-in and streamline recruitment efforts.

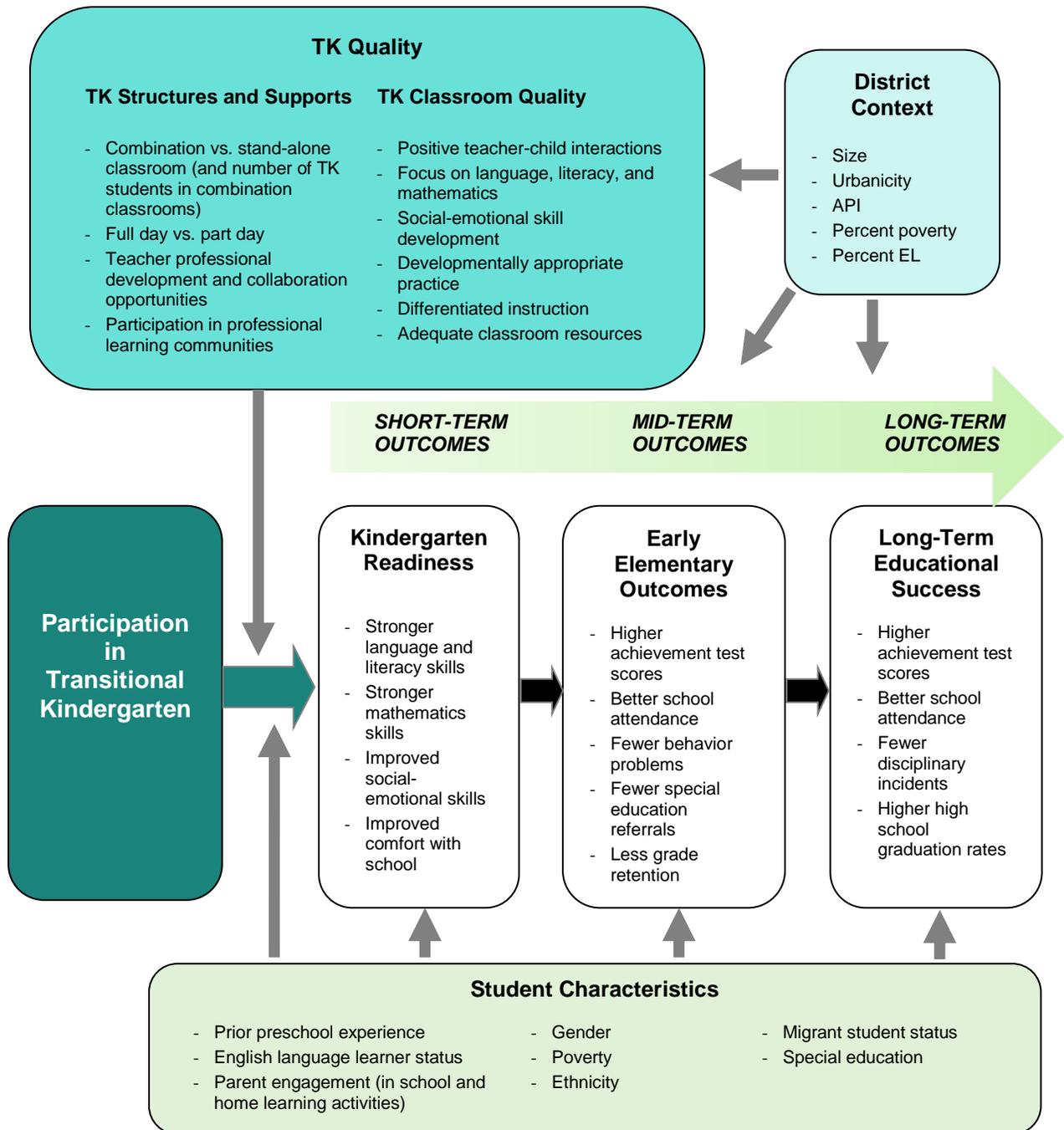
The design of the study is guided by a logic model and a set of research questions described in the next sections.

The TK Logic Model

TK is intended to prepare California's youngest students for kindergarten and for later school success. As shown in the logic model in Exhibit 1, TK is expected to have an impact on students' short-term, mid-term, and long-term outcomes. The short-term impacts of TK are hypothesized to manifest themselves in

increased kindergarten readiness in the domains of early language and literacy, mathematics, and social-emotional skills. Mid-term outcomes may include higher standardized test scores, improved school attendance, fewer behavior problems, fewer special education referrals, and less grade retention. As students experience increased educational success, we may ultimately expect to see higher graduation rates and fewer disciplinary incidents among the students who attended TK as young 5-year-olds.

Exhibit 1. TK Logic Model



The logic model highlights the importance of context as well, specifically the importance of quality indicators as moderators of the impact of TK on outcomes for students. TK quality can be divided into two domains: (1) structures, such as part-day or full-day options or combination or stand-alone classrooms, and supports available to teachers; and (2) quality at the classroom level—for example, teachers' instructional practices, their interactions with the students, and their facility with differentiating instruction for learners with different levels of need. Classrooms will vary along these dimensions, and higher quality TK classrooms are expected to support greater impacts for students.

District context also matters. With limited resources and few eligible students, small, rural districts are constrained in terms of how they can structure their TK programs. Large, urban districts with high levels of poverty and numerous English learners experience different challenges with their TK programs.

And, of course, students bring with them their own contexts and their own levels of readiness to learn. Limited or no prior preschool experience puts students at a disadvantage at kindergarten entry; attending TK may be especially beneficial for these students, boosting their school readiness skills and smoothing the transition to kindergarten. English learners are also a vulnerable group of students who may benefit from the additional year of language development to ensure that they are ready to engage in the exchange of academic language and take on the rigors of today's kindergarten classrooms.

Although this study will not address all questions suggested by the logic model, we will use the model as an initial guide for the study design and analysis.

Research Questions

The study will be guided by three sets of research questions. The first set focuses on **impacts of TK**:

1. Does TK participation improve kindergarten readiness in the domains of early literacy/language, mathematics, and social emotional skills?
2. To what extent are the impacts of TK sustained through the end of kindergarten?
3. Does the impact of TK change across cohorts?
4. How do student outcomes vary by child background characteristics, such as gender, English learner status, or poverty status?
5. How do student outcomes vary by district characteristics, such as district size or urbanicity?

The second set of questions addresses the **quality of TK** and how quality is related to student outcomes:

6. How do TK programs fare on measures of classroom quality?
7. How does TK quality change across cohorts?
8. How do student outcomes vary with TK classroom quality?
9. How do student outcomes vary by program structure, such as stand-alone versus combination classrooms?

Finally, we include an additional research question about **longer term impacts of TK**, for consideration at a later phase of the study:

10. Are the impacts of TK sustained through the early elementary school years?

This final question is not addressed by the proposed study design but could be incorporated into a second phase of the quality and outcomes study, after initial results are available, if this is of interest to the funders.

Proposed Methods and Analysis

The proposed study relies on an RD approach to understanding program impact. We also propose to examine the relationship between classroom quality and outcomes for TK students, how student outcomes vary by subgroup, and whether outcomes differ based on classroom structure (e.g., stand-alone versus combination classes) and other contextual (e.g., urbanicity, district size) variables. Our approach is described in further detail below.

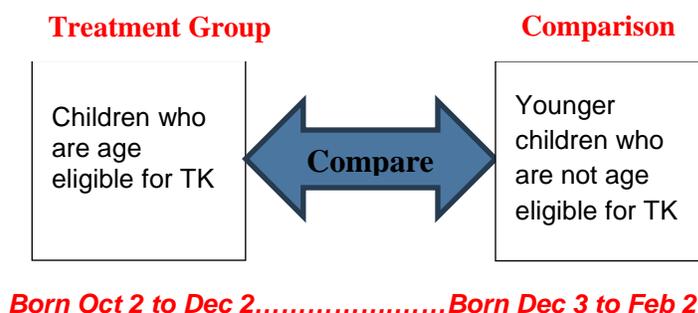
Identification Strategy: Regression Discontinuity

To measure the effect of TK, relative to “business as usual” (how similarly aged children would have progressed without the additional year of education), we would ideally randomly assign children to be either in TK or to continue with “business as usual,” which could include a range of arrangements, including child care, public or private preschool, Head Start, or remaining at home. However, random assignment would be difficult to defend and implement, and it would produce results that are not necessarily generalizable to the full population of TK-eligible children, because results would be limited to children whose parents would be comfortable with the uncertainty inherent in a randomized controlled trial setting. Fortunately for the study, eligibility for TK is limited to children in a very specific age range, which means that an RD design can be used to approximate the rigor and credibility of random assignment without actually randomly assigning children.

Previous studies have effectively used the RD method to study the impacts of early childhood education policies. Gormley and Gayer (2005), for example, evaluated Oklahoma’s prekindergarten program using the RD method and found that prekindergarten participants had stronger scores on assessments of cognitive, language, and motor skills. McEwan and Shapiro (2008) studied the effect of delayed primary school enrollment, comparing children whose birth dates were just above and below the cutoff date for first grade in Chile. Findings indicated that delayed primary school entry was associated with decreases in grade retention and increases in fourth- and eighth-grade test scores.

The RD study will compare children on either side of the December 2 cutoff date for TK eligibility, as shown in Exhibit 2. Children born December 2 or earlier, who are age eligible for TK, serve as the treatment group. Children who are too young to qualify for TK (i.e., those born December 3 or later) are the comparison group. These younger children will enter kindergarten at the same time as the TK children but without the TK experience.

Exhibit 2. The RD Approach



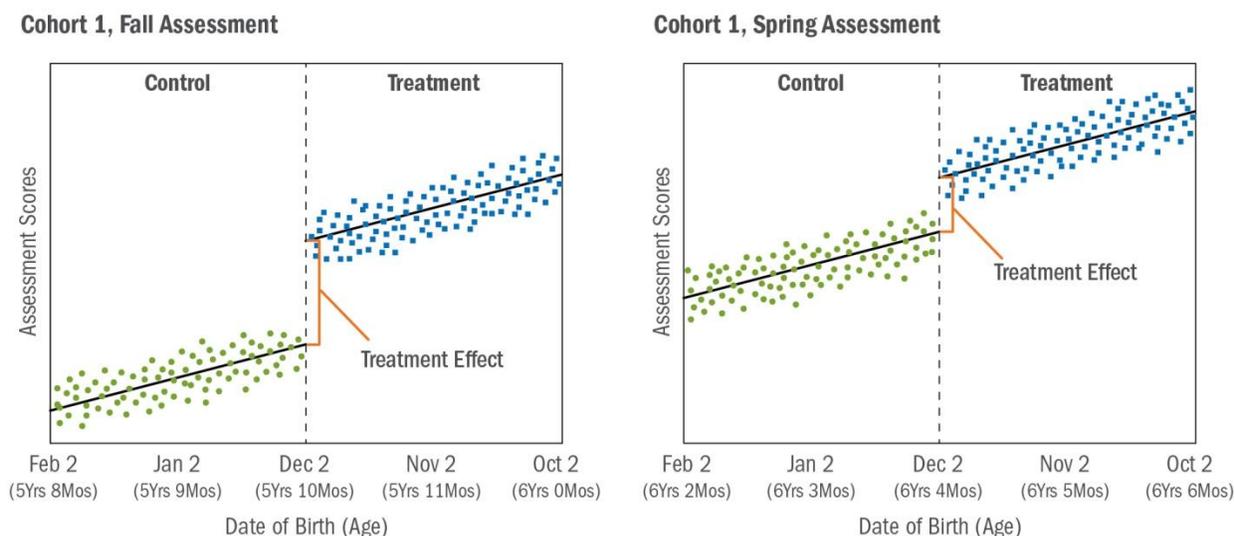
To estimate the treatment effect, we will identify children on both sides of the cutoff date and compare their outcomes. Our approach assumes that children’s eligibility for the program is entirely determined by their birth dates, which are distributed uniformly around the TK cutoff date. Because birth dates can be

assumed to be distributed randomly, the comparison between the outcomes of children with birth dates before and after the cutoff date can be likened to a randomized experiment after the children's ages (in days) are controlled for in the analysis.

The theoretical scatterplots shown in Exhibit 3 below represent potential findings from a regression discontinuity analysis. The x axes in both graphs represent children's birthdates with a 2 month window on either side of the discontinuity at December 2, the enrollment cutoff for TK, which is represented by a vertical dotted line. The x axis starts at February 2 and goes to October 2, so that child age increases from left to right along the axis. The treatment group is on the right side of each scatterplot, and the control group is on the left. Each blue square dot represents a child who was eligible for TK, and each green round dot represents a control group child, i.e. a child who was too young to qualify for TK. Children's scores on assessments are on the y axes. As indicated on the graphs, the size of the discontinuity in the regression line at the December 2 cutoff is the size of the treatment effect.

The graph on the left shows expected results from the fall assessment for cohort one, and the graph on the right shows expected results from the spring assessments for cohort one. For these graphs, we assume the fall and spring assessments will be approximately six months apart, for example in October and April. Therefore, child age starts at 5 years, 8 months in the left graph, and child age starts at 6 years, 2 months in the graph on the right. We expect the treatment effect to be larger in the fall of kindergarten than in the spring. That is, we expect the relative advantage that treatment group children demonstrate at school entry to diminish somewhat by the spring, after the control group children have participated in nearly a full year of formal schooling. We expect similar results for cohort two, although we expect that the treatment effect will be greater in magnitude as district administrators, principals, and teachers develop better systems for administering and teaching this new grade.

Exhibit 3. Visual Representation of the "Discontinuity" Anticipated for Cohort 1 at Fall and Spring Assessment Points



Sample, Data Collection, and Measures

The proposed approach to sampling and data collection is described below, as are the child outcomes and classroom quality measures that we recommend using for the study. After drawing a nested sample of

districts, schools, and children, the study team proposes to collect primary data on student outcomes and TK classroom quality. Direct assessments will measure children's early literacy and language, mathematics, and social-emotional skills, using assessment tools that have demonstrated predictive validity in previous studies. TK classroom observations and teacher surveys will focus on instructional quality, teacher-child interactions, student engagement, and time on various content areas. Kindergarten teacher and parent surveys will provide data on potential moderators, and a district survey will provide information about district TK policies and approaches as the programs become fully implemented.

We will incorporate two cohorts of children into the study. The first cohort will be made up of children entering kindergarten in 2014-15, half of whom will have been eligible for TK in the year before, and half of whom will not have been eligible and attended "business as usual" instead—in other words, a variety of pre-kindergarten experiences likely to include staying home, attending preschool part- or full-time, or being cared for by a family member or friend. The second cohort will be made up of children entering kindergarten in 2015-16, again with half having been eligible for TK and half having not been eligible for TK the prior year. Students in each cohort who attend TK will have their classrooms observed and their teachers surveyed during the TK year (spring of 2014 for Cohort 1 and spring of 2015 for Cohort 2) to understand quality and implementation of the program. There are two benefits to including two cohorts in the study: 1) to examine changes in TK quality over the two years, as the program becomes fully rolled out, and 2) to have additional data after two years so that we may investigate the impacts of TK for smaller subgroups of students (e.g., English learners) with sufficient power.

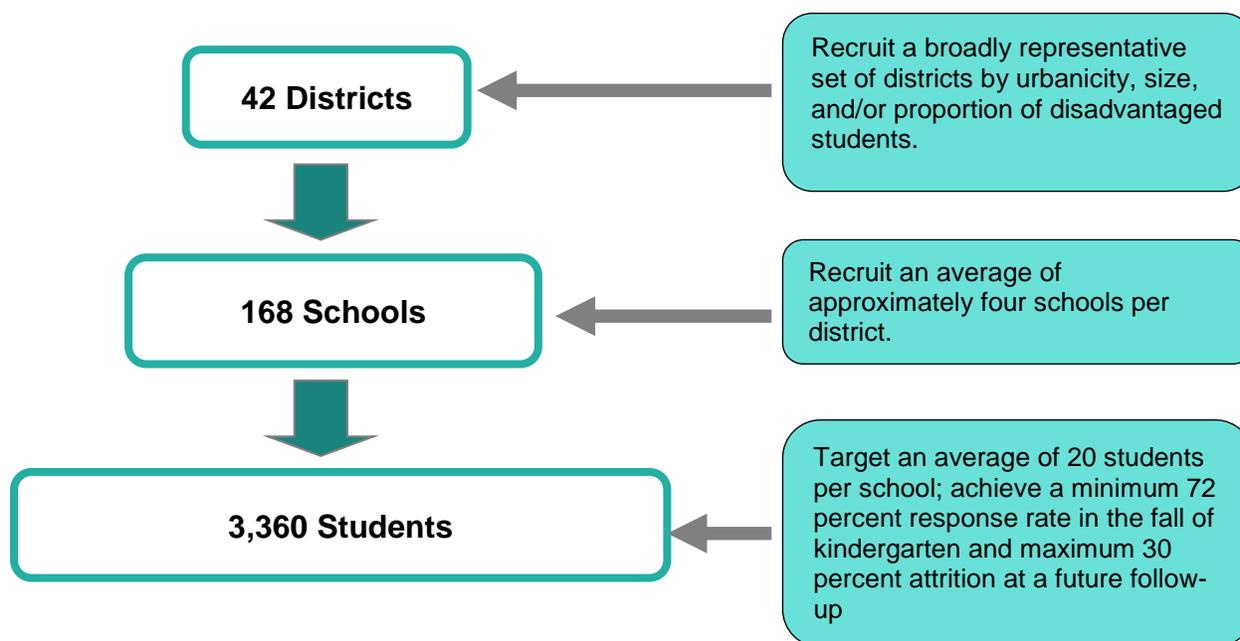
Sampling Plan

RD designs are less statistically efficient than a randomized assignment design (Schochet, 2008) because of the correlation between the treatment variable (eligibility for TK) and the "forcing" variable (age in days). Therefore, large sample sizes are needed to detect program effects. For the purpose of estimating the statistical power of this study, we assume a symmetrical four-month window of birth dates around the cutoff date for TK eligibility.¹ We also assume a minimum 72 percent response rate at the child level in the first wave of data collection and allow for attrition of up to an additional 30 percent of the sample at a potential follow-up data collection point (e.g., at second grade) so that the study is adequately powered for long-term follow-up. The study also is powered for intent-to-treat estimates of program impact, which provide conservative, policy-relevant estimates of treatment effects. Finally, the study is designed to allow for estimation of the impact of TK for subgroups of interest, such as English learner students, or students who receive free or reduced-price lunch.² Based on these parameters, our initial target sample consists of approximately 3,360 students nested within 168 schools and 42 school districts, as shown in Exhibit 4.³

¹ The optimal window around the enrollment cut, called the *optimal bandwidth*, is dependent on the distribution of scores on the outcome measures and cannot be determined in advance of data collection. We will collect data on all children eligible for the study and then use a statistical procedure to establish the optimal bandwidth around the cutoff date to maximize the analyses' statistical power. If using a narrower bandwidth would be more efficient, such an adjustment would be easy to implement and would result in more statistical power. In any case, we will run full sample impacts (ignoring the empirically established optimal bandwidth) as a sensitivity test to ensure that these statistical data manipulations do not have undue effects on the study's point estimates.

² Subgroup analyses will be conducted using combined (or "stacked") data from both cohorts.

³ The power analysis assumes the following: alpha = .05; a two-tailed test; power = .80; ICC = .15; treatment effect heterogeneity = 0; proportion of students who receive TK = 50 percent (symmetrical RD design); R² Level 1 = .2; R² Level 2 = .1; no. of Level 2 covariates = 0; design effect = 4; correlation between treatment and birth date = .8. The design effect was selected based on guidance in Schochet (2008).

Exhibit 4. District, School, and Student Sampling Design

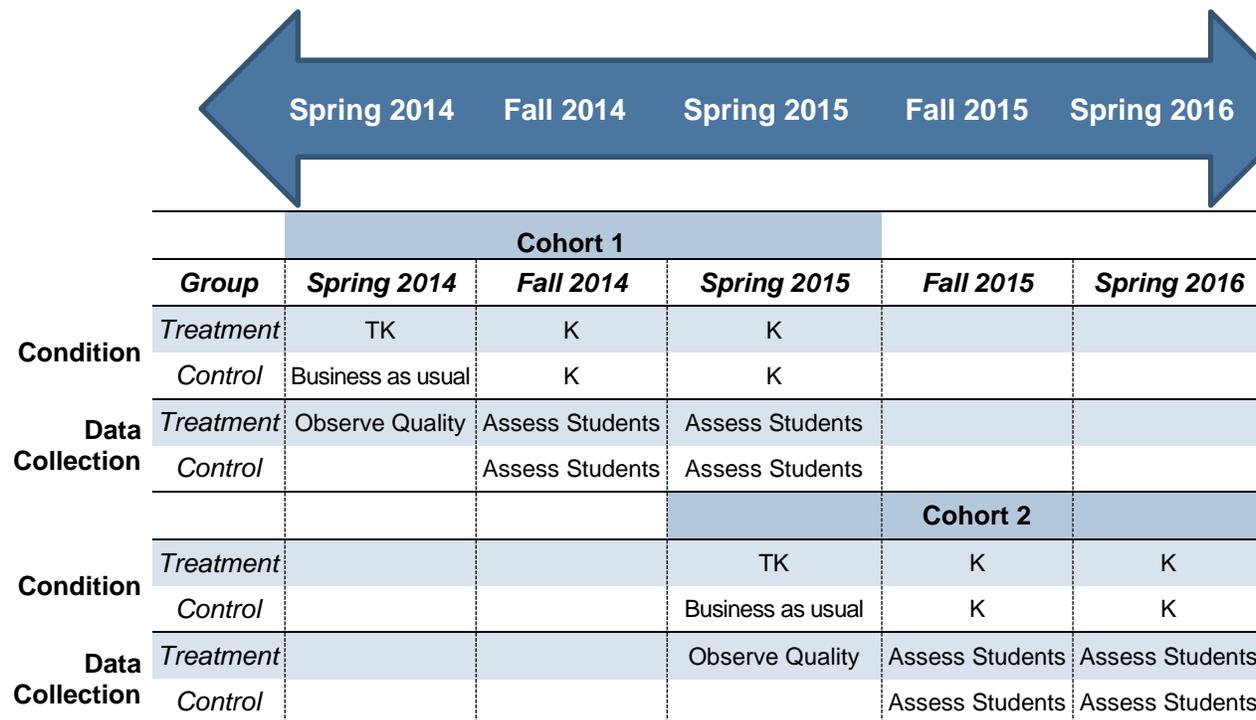
Using this sampling design, the research team will draw a broadly representative sample of districts, ensuring that the sample represents variation in district characteristics, such as size, urbanicity, and/or measures of student disadvantage, such as the proportion of students who are English learners and the proportion of students eligible for free or reduced-price lunch. All or most of the schools offering TK in small- to medium-sized districts will be invited to participate in the study, while a subsample of schools within larger school districts will be included. When selecting schools, the study team will consider total TK enrollment and whether the school is a hub. Within each school, all students who fall within the birth date eligibility window for the study will be part of the study.⁴ We will abide by all institutional review board (IRB) requirements regarding contacting participants and obtaining consent.

Data Collection and Measures

The study team will conduct direct assessments of students and observe quality of TK classrooms. Primary data collection will allow the study team to gather high-quality and uniform student outcome data in a variety of domains with normative measures that capture the full variation of children's knowledge and skills. This approach also will allow us to easily link data on TK classroom quality to child outcomes. Two observers will conduct half-day observations of TK classroom quality in the spring of the TK year for each cohort. Child assessments will take place in one-on-one sessions of approximately 45 minutes during the following fall and spring, when the former TK students and their peers enter kindergarten. Kindergarten teachers also will be invited to complete assessments of students' social-emotional skills and classroom behavior. The data collection timeline is shown below in Exhibit 5

⁴ Children with birth dates plus or minus two months of the December enrollment cutoff will be part of the study in both Cohort 1 and Cohort 2.

Exhibit 5. Data Collection Timeline



* The treatment group includes children born October 2 to December 2. The control group includes children born December 3 to February 2.

Child Outcome Measures. The proposed study will measure children’s competence in the domains of early language and literacy, mathematics, and social-emotional skills. We present our proposed set of options for measures below and summarized in Exhibit 6. The study team will present final recommendations on child outcome measures to the funders and TAG in the fall in the form of a design memo. The recommendations will be based on a review of the literature and consideration of the following criteria:

- Likelihood that TK would impact the constructs measured by the assessments, based on the logic model
- Extent to which the assessments are associated with future academic achievement
- Administration time
- Ease of administration for non-clinicians
- Availability of the assessments in Spanish (or the feasibility of translation)
- Ability to discriminate between children at different ability levels.
- Extent to which the assessments tap conceptual knowledge, rather than just rote knowledge

First, we will screen study children to determine if they are proficient enough in English to be assessed in English, using the Pre-LAS Art Show and Simon Says subtests. These subtests were used in the Early Childhood Longitudinal Study, Kindergarten Class of 1998 (ECLS-K) as well as in the Early Childhood

Longitudinal Study, Birth Cohort (ECLS-B). If study children are Spanish speakers and are not proficient enough in English to be assessed in English, they will be assessed in Spanish.⁵

Based on feedback from the TAG and our preliminary research, we propose to assess language and literacy skills using measures that evaluate receptive and expressive vocabulary, phonological awareness, and letter and word recognition. These language and early literacy skills are recommended for assessment because they are associated with later reading success and because they are skills likely taught in TK and kindergarten. Consequently, we considered instruments that measure these skills and evaluated them in terms of age-appropriateness, length, appropriateness for a diverse group of children, and psychometric properties. Specifically, we recommend using a combination of the Comprehensive Test of Phonological Processing (CTOPP) elision scale, and the Woodcock-Johnson Letter Word Identification test. Together, these measures will take approximately 15 minutes.

To assess early mathematics skills, we propose to use either the Woodcock Johnson Test of Achievement, Applied Problems and Quantitative Concepts subtests, or the Research-based Early Mathematics Assessment (REMA) short form. Both of these tests have many advantages, including good psychometric properties and short administration times. Additionally, both options will assess a range of mathematical skills, and not just numeracy. The REMA short form is a relatively new instrument, but was developed by top researchers in the field and is designed to reflect the natural progression of mathematics skill development as uncovered through recent research. However, if we use the Woodcock-Johnson tests, we will be able to compare our results to those in other large studies, given how commonly these instruments are used in early childhood research. We had considered the Test of Early Mathematics Ability, which assesses a broad range of skills and captures conceptual knowledge, but the administration time is too long. If a custom mathematics assessment is desired in place of these two options, AIR will provide a separate addendum to this proposal to develop such an assessment and conduct a validity study.

We will assess several aspects of children's social-emotional development. First, executive function will be measured through direct assessment. Executive function enables children to inhibit or ignore distractions in order to focus on the task at hand. Additionally, executive function includes being able to be cognitively flexible and change one's understanding and approach to a problem in pursuit of the answer. The last main component of executive function is working memory, which allows students to keep in mind the aspects of a problem needed to find a solution or to understand a new concept. It is important to measure executive function in our study children, because it is important for school success. Thus, we propose to include a direct assessment of executive function, such as the Head, Toes, Knees and Shoulders Task, which has been predictive of future academic achievement in several studies (e.g., Blair & Razza, 2007; McClelland et al., 2007; Ponitz, McClelland, Matthews, & Morrison, 2009).

We also will include a teacher report of children's classroom behaviors. Options include the Child Behavior Scale, Preschool Kindergarten Behavior Scale, or select scales from the updated version of the Social Skills Rating System, called the Social Skills Improvement System Rating Scales. The Child Behavior Scale measures children's aggressive, withdrawn, and prosocial behaviors with peers. These behaviors are important to measure, because children who are aggressive or asocial are more likely to have later adjustment problems (Ladd & Profilet, 1996). Also, children's prosocial behaviors in early childhood are predictive of later academic achievement (Caprara et al., 2000). The Preschool Kindergarten Behavior Scale and the Social Skills Improvement Rating Scales both measure social skills and problem behaviors, such as externalizing and internalizing behaviors. The Social Skills Improvement Rating Scales also measures academic competence, based on teacher report.

⁵ Note that we will not be able to assess children in any language other than English or Spanish, although consent materials will be available for parents in additional languages as necessary.

Exhibit 6. Recommended Child Assessments

Measure	Constructs measured	Time required
Pre-LAS Simon Says and Art Show	Language Screener (Receptive and Expressive Language)	5-10 minutes
Comprehensive Test of Phonological Processing (CTOPP) Elision scale	Phonological Awareness	10-15 minutes
Woodcock-Johnson Letter and Word Recognition	Letter and Word Recognition	
Research-based Early Mathematics Assessment (REMA) short form <i>or</i> Woodcock-Johnson Applied Problems and Quantitative Concepts subtest <i>or</i> newly-designed subset of items from the REMA or Test of Early Mathematics Ability (TEMA), to be piloted ⁶	Mathematics	15-20 minutes
Head, Toes, Knees, Shoulders	Executive Functioning	5 minutes
TOTAL		35-50 minutes

English learners (ELs) are a substantial population in California, and it is important for measures to have been shown to be valid with ELs, and to be culturally unbiased. All of the recommended measures presented above are available in Spanish, have been validated with Spanish speakers, and have no associated cultural bias concerns.

TK Classroom Observations. Classroom observers will gather data on classroom quality during the TK year. We propose to include the CLASS instrument, which measures the extent to which the teacher provides emotional support, classroom organization, and instructional support. We will supplement the CLASS with one or two additional measures that assess other domains of quality, such as the quality and quantity of instruction in different content areas, the extent to which children are engaged, teacher sensitivity, and the availability of learning materials. (See Exhibit A.2 in the appendix for more details on the other measures under consideration, including the Classroom Observation of the Early Mathematics Environment and Teaching [COEMET], Early Language and Literacy Classroom Observation [ELLCO], Emerging Academics Snapshot, and Teacher Behavior Rating Scale [TBRSS]).

TK Teacher Survey. A survey will be administered to TK teachers in the spring to gather data on teacher background characteristics, classroom structure (e.g., stand-alone versus combination, full day versus half day), instructional practices (including grouping strategies and time spent on various topics), use of assessments, and strategies and opportunities for parent engagement, using the same questions used on the teacher survey for the implementation study, where appropriate.

⁶ A separate proposal addendum will be submitted for such a validity study, should this option be exercised.

Kindergarten Teacher Survey. A kindergarten teacher survey will provide data on teacher background characteristics and classroom structure variables, such as full day versus part day. We also will ask kindergarten teachers to rate their students, using one or more teacher report measures, as described previously.

Parent Survey. In the process of obtaining consent from parents (or in a follow-up communication with parents), we will ask three *brief* questions about the child's experiences prior to kindergarten: what care arrangement the child was in prior to kindergarten, for how long the child participated in this care arrangement, and for how many hours per week.

District Survey and Interview. Because 2013–14 is the second year of TK in California, with full implementation in 2014–15, it is expected that the program will continue to change and develop throughout the course of those two years. How the program is implemented statewide and by district is critical context to understand the emerging child outcomes being measured in this study. Thus, we will ask TK administrators in each sampled district to complete a brief survey each year. This information will be used to present descriptive information about TK policies and approaches and supports for teachers and schools as TK evolves. We also will explore how districts incorporate migrant students and special education students in TK programs and perceptions of impacts through district interviews, because these groups will likely not be large enough to include in the main set of outcome analyses.

Analytic Approach

The study team will use two approaches to model the impacts of TK: fuzzy RD estimates that account for noncompliance with enrollment guidelines and sharp RD estimates that show the effect of offering the TK program regardless of whether children participate. In combination, these two approaches to statistical modeling will provide upper and lower bound estimates of the causal effect of TK attendance on child outcomes.

Noncompliance with enrollment guidelines leads to “fuzziness” at the December 2 enrollment cutoff, where the effect of the discontinuity (i.e., TK) is to be estimated. Because compliance with the eligibility cutoff date is likely to be correlated with other background characteristics such as race or ethnicity and family income, excluding noncompliers from the analyses would introduce selection bias. We will use a fuzzy RD design to eliminate this source of selection bias.⁷ The fuzzy RD design employs a two-stage least squares correction (2SLS) to account for noncompliance with the cutoff date and will allow us to estimate the effect of the treatment on those children who received it by using predicted enrollment, rather than actual enrollment, as the primary explanatory variable in the impact model.

The first stage of the fuzzy RD design still compares children who are age eligible for TK to those children who are not eligible. Exhibit 7 illustrates this comparison. The treatment group will include the compliers and noncompliers shown in Cells A and B; children who were eligible but whose parents chose not to enroll them in TK will be included in the treatment group. The comparison group will include the noncompliers and compliers shown in Cells C and D; children who participated in TK even though they were not age eligible will be included in this group.

⁷ See Hahn, Todd, and Van der Klaauw (2001), Jacob, Zhu, Somers, and Bloom. (2012), or Imbens and Lemieux (2008) for a more detailed discussion of the fuzzy RD method.

Exhibit 7. Framework for a Fuzzy RD Study

		Enrolled in TK	
		Yes	No
Eligible for TK	Yes (Treatment Group)	Complier A	Noncomplier B
	No (Comparison Group)	Noncomplier C	Complier D

In the first-stage model for the fuzzy RD, the probability of receiving the treatment is estimated using the treatment assignment and forcing variables, where ASSIGN is the treatment assignment (i.e., whether the child is or is not eligible for TK based on his or her birth date), AGE is the child’s age measured in days, and TREAT is whether the child attended TK. In the second-stage model, the estimate from the first-stage model is used as a predictor variable, where TREAT_{hat} is the estimated likelihood of treatment from the first stage model, AGE is the child’s age measured in days centered at the cutoff, and COV is a vector of covariates. Random error terms “e”, at the student level, and “u” at the school level are assumed to be identically and independently distributed.

$$\text{First-stage model: } TREAT = \beta_0 + \beta_1 \text{ ASSIGN} + \beta_2 \text{ AGE} + \beta_3 \text{ COV} + e$$

$$\text{Second-stage model: } Y = \beta_0 + \beta_1 \text{ TREAT}_{\text{hat}} + \beta_2 \text{ AGE} + \beta_3 \text{ COV} + u + e$$

In addition to estimating fuzzy RD estimates in a 2SLS framework, we will also estimate so-called “intent to treat” (ITT) effects at the cut point, using sharp RD analyses, where we have only the second stage model and TREAT is used instead of TREAT_{hat}. These estimates capture the effect of offering a parent the opportunity to place his or her child in TK, regardless of whether the parent takes up that offer. These ITT estimates provide a useful sensitivity test for the fuzzy RD estimates because both are expected to move in the same direction, even if the fuzzy RD estimates may be larger in an absolute sense. ITT also provides a more conservative estimate of program effects, which is very policy relevant.

Estimating the Impact of TK

Our first set of analyses will utilize the two RD approaches described previously (fuzzy and sharp RD) to address research questions related to the impact of TK: (1) Does TK participation improve kindergarten readiness in the domains of early literacy/language, mathematics, and social emotional skills? (2) To what extent are the impacts of TK sustained through the end of kindergarten? (3) Does the impact of TK change across cohorts? We also will explore outcomes by contextual variables: (4) How do student outcomes vary by child background characteristics, such as gender, English learner status, or poverty status? (5) How do student outcomes vary by district characteristics, such as district size or urbanicity? The analytic approach to addressing each research question is described below and summarized in Exhibit 8.

TK Impacts at Kindergarten Entry. To answer our first research question, the study team will conduct a series of RD models examining students’ early language and literacy, mathematics, and social-emotional skills, as measured in the fall of kindergarten. A full set of analyses will be conducted using

both estimation approaches described previously: fuzzy and sharp RD. More detailed analyses also will be conducted on select scales from the child assessments. For example, in the social-emotional domain, we may examine children's executive functioning separately from their pro-social behaviors and classroom self-regulation.

Sustained Effects of TK. RD models will be repeated with spring kindergarten measures to ascertain the extent to which effects of TK evident at kindergarten entry persist through the end of the kindergarten year. These analyses will address questions about the persistence of any relative advantage that TK students exhibit at kindergarten entry, albeit during a relatively short time frame.

Variation in Student Outcomes, by Student or District Characteristics. To further explore associations between TK participation and child outcomes, the study team will run models on subgroups defined by student and district characteristics. Examples of student-level moderators include English learner status, gender, eligibility for free or reduced-price lunch, and preschool participation. We also may examine subgroups defined by district characteristics such as district size and urban versus rural school districts. To increase the likelihood that statistical models have adequate statistical power to detect program effects, subgroup analyses will be conducted with stacked data from both cohorts after Cohort 2 data collection is complete and will focus on subgroups that comprise a sizable proportion of the sample.⁸

Changes in TK Impact. We hypothesize that the impact of TK will increase as the program is fully implemented. We will compare impact estimates from the fall of kindergarten for Cohorts 1 and 2 as well as impact estimates from the spring of kindergarten for both cohorts to examine this hypothesis.

Assessing the Relation of TK Quality and Structure to Outcomes

The second set of analyses will address our research questions about the **quality of TK**: (6) How do TK programs fare on measures of classroom quality? (7) How does TK quality change across cohorts? (8) How do student outcomes vary with TK classroom quality? (9) How do student outcomes vary by program structure, such as stand-alone versus combination classrooms? Analyses will include descriptive statistics and regression analyses.

TK Quality. We will begin our quality analyses by examining scores on observational measures of classroom quality, including total scores and scale scores. For example, we will examine total CLASS scores as well as domain scores on emotional support, classroom organization, and instructional support. Results will be presented as descriptive statistics, including the mean, median, standard deviation, and range.

Changes in TK Quality. We hypothesize that the quality of TK will increase as the program is fully implemented. To examine this hypothesis, we will prepare descriptive statistics for Cohort 1 and Cohort 2, which will allow for a comparison of overall program quality in Years 2 and 3 of program implementation. We also may compare within-teacher changes in quality, with regression models that use teacher background characteristics and/or classroom structure variables as predictors.

Variation in Outcomes by TK Quality and Program Structure. We will address research questions about program quality and structure using three approaches. First, classroom quality will be treated as a potential moderator of associations between TK attendance and children's skills at kindergarten entry. These regression models will test whether children who attended higher quality TK classrooms display greater advantages in early literacy, mathematics, and social-emotional skills at kindergarten entry than children who attended lower quality TK classrooms. These analyses will be designed to be quite specific

⁸ The study will not be adequately powered to examine very small subgroups.

about the domains of quality expected to be associated with particular child outcomes. For example, do children who attend TK classrooms taught by teachers with better behavior management display better self-regulation skills at kindergarten entry? Do children who receive more mathematics instruction in TK score higher on a measure of conceptual mathematics knowledge? Second, we will examine the extent to which student outcomes vary by program structure variables, such as stand-alone versus combination classrooms. Third, we will treat classroom quality as a moderator that may help explain associations between TK attendance and children’s skills at kindergarten entry. These models will be of particular interest if we find that both TK quality and TK impacts increase with full implementation.

Exhibit 8. Statistical Model and Analytic Approach to Addressing Each Research Question

Research Question	Statistical Model	Analytic Approach/Data Source
1. Does TK participation improve kindergarten readiness in the domains of early literacy/language, mathematics, and social emotional skills?	Regression discontinuity models*	Run separate models on: Cohort 1, Fall Kindergarten Assessment Cohort 2, Fall Kindergarten Assessment
2. To what extent are the impacts of TK sustained through the end of kindergarten?	Regression discontinuity models	Run separate models on: Cohort 1, Spring Kindergarten Assessment Cohort 2, Spring Kindergarten Assessment
3. Does the impact of TK change across cohorts?	Wald test or Lagrange multiplier test of statistical significance of variation in TK effect estimates across cohorts	Conduct two separate comparisons of Cohorts 1 & 2: Cohort 1 vs. Cohort 2, Fall Kindergarten Assessment Cohort 1 vs. Cohort 2, Spring Kindergarten Assessment
4. How do student outcomes vary by child background characteristics, such as gender, English learner status, or poverty status?	Regression Discontinuity Models on Subgroups	Run models on combined data from Cohort 1 and Cohort 2, Fall Kindergarten Assessment
5. How do student outcomes vary by district characteristics, such as district size or urbanicity?	Regression Discontinuity Models on Subgroups	Run models on combined data from Cohort 1 and Cohort 2, Fall Kindergarten Assessment
6. How do TK programs fare on measures of classroom quality?	Descriptive statistics	Calculate separate statistics for: Cohort 1 Spring TK Classroom Observation Cohort 2 Spring TK Classroom Observation
7. How does TK quality change across cohorts?	T-tests, OLS regressions	Compare Cohort 1 Spring TK Classroom Observation to Cohort 2 Spring TK Classroom Observation
8. How do student outcomes vary with TK classroom quality?	OLS regressions (moderation analysis)	Run separate models for: Cohort 1 Spring TK Classroom Observation Cohort 2 Spring TK Classroom Observation
9. How do student outcomes vary by program structure, such as stand-alone versus combination classrooms?	OLS regressions (moderation analysis)	Run separate models for: Cohort 1 Spring TK Classroom Observation Cohort 2 Spring TK Classroom Observation

Workplan and Timeline

This section describes our proposed workplan and timeline for the evaluation. The study includes 11 tasks:

- Task 1: Planning, Preparation, and Training
- Task 2: Sampling and Collecting Student Record Data
- Task 3: Recruitment
- Task 4: TK Classroom Observations
- Task 5: TK Teacher Survey
- Task 6: District Survey and Follow-Up Interviews
- Task 7: Kindergarten Student Assessments
- Task 8: Kindergarten Teacher Survey
- Task 9: Data Analysis
- Task 10: Reporting
- Task 11: Project Management

Task 1: Planning, Preparation, and Training

The first task will be to refine and finalize our data collection plan to guide the data collection activities for the duration of the study.

Task 1.1: Conduct Key Informant Interviews

To ensure the success of the study, we will need to work closely with districts to determine the best strategies for recruiting schools and families. We anticipate that we will need to employ multiple strategies for recruitment and that districts will have different requirements, preferences, and recommendations for how we approach schools and families. To explore the range of options, we will reach out to several large districts at the beginning of the study to introduce the study to them and learn more about their interests in the topic; what might engage the district, school staff, and families in terms of incentives; and what district-specific data collection guidelines we will need to follow. We worked with many districts during the implementation study and have developed relationships that will facilitate these discussions. From these discussions, we will refine and finalize our data collection plan, including recruitment plans, incentives, and procedures for obtaining the needed data.

Task 1.2: Develop Recruitment Materials

Second, we will develop recruitment materials for districts, schools, teachers, and families. This will include, for example, the following:

- Letters, e-mails, and phone scripts for contacting district and school administrators
- A brief study overview to share with districts, schools, and teachers, presenting the importance of the study, the relevance of the research to their practices, and information about anticipated frequently asked questions
- A brief report of implementation study findings, by district type (e.g., small, medium, large)

- Memorandum of understanding and data-sharing agreement templates for use with districts to enable the acquisition of data, such as student rosters and directory information (e.g., date of birth and parent contact information)
- Scheduling guidelines for school liaisons
- Study brochure or flyer for parents, featuring the importance of their involvement in this research, the value of this study in terms of the education of their child, and responses to questions parents might have (in English and Spanish)
- Parent consent forms (in English and Spanish)
- Assent scripts for student assessments
- Training guide for staff recruiters

We also will develop a database for documenting recruitment and managing all data collection activities in an organized and efficient way.

Task 1.3: Finalize Measure Selection

Third, we will finalize measure selection decisions. We have listed our recommended measures to include for assessing outcomes and quality in this proposal, but we will finalize the selection of these tools in the fall. We will have a follow-up discussion with Margaret Burchinal of Frank Porter Graham Child Development Institute, who made many recommendations during our TAG meeting last spring and who offered to provide additional guidance on this topic. We will prepare a memo that summarizes our selection decisions and share the memo with the funders and the TAG for comment and final approval.

Task 1.4: Obtain IRB Approval

After measures are selected, the plan for data collection is finalized, and the recruitment materials are developed, we will submit the entire research protocol to AIR's IRB for approval. AIR's IRB, which is registered with the Office of Human Research Protection and operates under a Federalwide Assurance comparable to many universities, is responsible for reviewing all research conducted by the organization and its subcontractors. The IRB ensures that projects involving human subjects comply with professional standards and government regulations designed to safeguard participants and that research team members, including subcontractor staff, are adequately trained. The criteria that a study must meet for AIR's IRB approval include assurances that risks to participants are minimized, risks are balanced by benefits, subject selection is equitable, participants are informed about risks and give uncoerced consent, privacy of the subjects is adequately protected, confidentiality of the data is adequately protected, and the rights and welfare of populations that might be vulnerable to coercion or undue influence are protected. All research team members participating in the study have been trained on human subjects protections and are well versed in procedures that relate to the responsible conduct of research. Data security is also a top priority at AIR, and all procedures for replacing personally identifiable information with identification numbers and storing data in secure, password-protected files are followed. After IRB approval is granted, we can begin recruitment. IRB at the school-district level is addressed in Task 3.

Task 1.5: Train Field Staff

An important part of the data collection effort is quality control, which starts by selecting qualified data collectors and providing them with comprehensive training. For example, AIR will provide a two-day CLASS training with a certified AIR CLASS trainer. Each data collector will be required to pass the online Teachstone CLASS reliability test following the training. Additional training also will be provided

on the other observation protocols and student assessment measures selected to ensure strict adherence to developers' standards and interrater reliability. The data collectors also will be given an overview of the study and will be trained on all research procedures, including scheduling, professional conduct during the observation, and documentation and record keeping.

Task 2: Sampling and Collecting Student Record Data

The second task will be to draw the sample of districts and schools in which data will be collected for Cohorts 1 and 2.

Task 2.1: Draw District Sample

Our sampling statistician will work with analysts on the project to draw a broadly representative sample designed to yield approximately 42 districts in the state per the sampling plan described in the study design. Backup districts will be identified for each sampling group to ensure that, if districts refuse to participate (after all routine efforts to include them are exhausted), we can quickly replace them with another district that is similar in characteristics (e.g., region, size, urbanicity, percentage English learner).

Task 2.2: Draw School Sample

From within the sampled districts, all schools offering TK will be identified. We will draw on information from the implementation study to identify schools with TK, and we will confirm school lists with districts during the initial recruitment phase (see Task 3). We will target all schools offering TK in small- to medium-sized districts for inclusion in the study and select a subsample of schools offering TK in larger school districts. Classroom observations in TK classrooms (see Task 4) will occur in these schools for both Cohort 1 (spring 2014) and Cohort 2 (spring 2015). Because some of these schools will be hubs, where TK students from across the district attend TK, we will follow the TK children, after their TK year, into the schools where they attend kindergarten. We anticipate the study will include approximately 168 schools.

Task 2.3: Identify Pool of Eligible Students from District Data

As described in the study design, we will assess students born between October 2 and December 2 (treatment students) and between December 3 and February 2 (control students) for Cohort 1.⁹ To identify the pool of students eligible for inclusion in the study, we will begin with TK students in the 168 schools in which TK classrooms are observed. Using student rosters obtained from the districts, we will identify all TK students in these schools who are in the age eligibility window (primarily October 2 to December 2). These students will be consented (through the parental consent process) in the spring and assessed in the fall. Most of these students will remain in the same school for kindergarten as their TK classrooms; however, some students will move to other schools in the district, especially those students in TK hub arrangements. We will follow these students to their kindergarten schools if the students remain within the same districts, and we will use student data provided by districts to identify additional kindergarten students, in those schools, who are in the age eligibility window (primarily December 3 to February 2). These students will be consented as early as data become available and will also be assessed in the fall.

⁹ The eligibility criteria for inclusion in the study will expand to include students born three months before and after the enrollment cutoff in Cohort 2. Specifically, children born in the 92-day period from September 2 through December 2 will be included in the treatment group, and the comparison group will include children born in the 92-day window from December 3 through March 4.

Task 3: Recruitment

Recruitment will be one of the most important tasks of the study. Without sufficient attention to this step, we will not have enough students to assess to sufficiently power the RD analysis.

Task 3.1: Recruit Districts

We will begin by reviewing district research applications and protocols to anticipate the issues that might be of concern to the district. Next, we will contact the TK administrator (or other contacts identified) at the district office to introduce the study and find out the best approach to gaining access to the district. We will complete research applications as necessary and provide districts with a copy of our research protocol and IRB determination form.

Achieving district buy-in will pave the way for school and parent recruitment, so this will be a top priority. We will use several strategies to garner district interest in the study:

- We will share results from the study in three forms:
 - District report of the implementation study findings (reported by district type: small, medium, large)
 - District reports of impact study findings, reported by district (to the extent that district Ns are high enough), and compared with overall study results
 - Research briefs shared broadly
- We will provide districts with a modest payment for providing data.
- We will be flexible in our approach to reaching out to parents, adhering to district requirements and preferences.

We will rely on an experienced team of recruiters to communicate with district administrators and convince them to participate in the study.

Task 3.2: Recruit Schools

After districts are on board, we will begin recruiting schools. After confirming the list of eligible schools, we will select an average of four schools per district.¹⁰ We will ask districts to reach out to schools and encourage them to participate. We found this strategy to be very helpful for achieving school participation in the implementation study. After districts have shared information about the study with the schools, we will contact the school administrators and invite them to participate. We will offer to share the same district-level reports with school administrators so that they can see the results of the study and how their district performs. After we have obtained cooperation from the school, we will identify a study liaison within the school. The study liaison will be compensated with an honorarium for his or her time. The study liaison will aid in terms of the logistics of conducting the study within the school, including facilitating access to teachers, scheduling observations, identifying space for the direct child assessments, and scheduling assessments. The study liaison could be an administrative staff member or some other qualified staff member within the school.

¹⁰ Districts with TK hubs will have fewer schools selected than expected, given the district's size, but additional schools will be added at the assessment phase as we track students into their home schools for kindergarten.

Task 3.3: Recruit and Consent Parents

After schools are recruited, we will work with each district and the school to coordinate access to parents to invite them to participate. Districts will have different requirements for informing parents about their child's participation in research, and we will follow district policies as appropriate. There are three potential options for obtaining parent permission for this study. We will seek approval for each from our IRB to allow us the flexibility to use whichever best matches district requirements:

1. **Obtain Implicit Consent.** For districts that allow implicit consent from parents, we will provide teachers with letters to parents, which inform parents about the study and give parents an option to opt out of the study, to send home with students. Parents who do not return the form indicating that they would not like their child to participate are assumed to give their permission.
2. **Obtain Explicit Consent, With No Direct Contact With Parents.** Some districts will require explicit consent but will not be willing to provide student rosters with parent contact information so that the study team can contact parents directly. In these scenarios, we will rely on teachers to obtain parent consent. As with Option 1, we will provide teachers with letters to parents to send home with students. However, unlike Option 1, these letters also will contain permission forms that will require parents' signatures in order to include the student in the study. These forms also can include some basic information about the child, such as prior preschool experience. Teachers will be given an incentive to distribute the letters and will be given a small incremental incentive for each consent form returned (signed or not).
3. **Obtain Explicit Consent, With Direct Contact With Parents.** A third scenario is one in which explicit consent is required, but the district is willing to share parent contact information. In these districts, we will first rely on teachers to obtain consent (as with Option 2), because they already have a relationship with parents; however, we will be able to follow up with nonresponsive parents, as needed, to ensure a high compliance rate.

These approaches will be used for treatment students in TK classrooms in spring 2014 (Cohort 1) and 2015 (Cohort 2). We also will use these strategies for control students in kindergarten classrooms in fall 2014 (Cohort 1) and 2015 (Cohort 2). We also will explore strategies for reaching parents of control students *before* they enter kindergarten, such as during kindergarten roundup activities, information nights, or other outreach efforts in the spring or summer prior to kindergarten entry.

Consent forms will be available in English, Spanish, and up to two other language as needed. Specific languages will be determined after initial conversations with districts (Task 1.1).

Consent forms for parents in districts requiring explicit consent will include a couple of questions about their child's prior preschool experience to use as a covariate in the analysis. We will collect this information in other ways for families in districts that allow for implicit consent. If the districts maintain this information in their district data system, we will request the data for all participating students. If the district does not maintain the data, we will contact the parents in a separate communication—either through a brief online survey, e-mail request, mail survey, or phone call—to ensure that we have complete information for a significant percentage of study students.

Task 4: TK Classroom Observations

As noted previously, we propose to observe TK classrooms using the CLASS observation tool, and, because the CLASS focuses on teacher-child interactions, the observation component may be supplemented with up to two additional measures focused on the content of instruction. After school administrators agree to allow their schools to participate in the study, we will reach out to TK teachers to

schedule and conduct these classroom observations. Classroom observations will be conducted in spring 2014 (Cohort 1) and 2015 (Cohort 2).

Task 4.1: Schedule Observations

Trained field staff will work with the study liaison, which could be the TK teacher, to schedule the observations. Prior to scheduling, a letter will be sent to each selected school, with an overview of the observation process, including what the process involves, observation procedures, length of time, and related information. We also will provide answers to frequently asked questions and contact information for key study staff. The observer will call each school (documenting all contacts) and coordinate the scheduling of the observations.

Task 4.2: Conduct Observations

Each selected TK classroom will participate in one observation that will last approximately three to four hours (assuming two observers). One observer will use the CLASS observation tool; the other observer will use one or two supplemental measures to capture the content of instruction and the classroom structure. Observers will record their observations and coding on forms that will be entered into a database for analysis.

AIR will establish an observation team lead, who will closely monitor the quality of the observation data. Observers also will provide weekly updates related to progress to the observation team lead. To ensure quality control, in up to 20 percent of the observations, we will send a second observer to co-observe with the classroom's assigned observer to compare scores to ensure the observer's reliability on the instrument. These quality assurance visits will occur throughout the project to ensure ongoing reliability.

Task 5: TK Teacher Survey

To supplement the classroom observations and to collect additional information about teacher qualifications and experience, we will administer a survey to all TK teachers in the study.

Task 5.1: Develop Survey

The TK teacher survey will be based on the teacher survey developed for the implementation study. Select items from the comprehensive implementation study survey will be compiled to create a shorter survey that still captures a breadth of information to characterize the experience of the TK students. Items will include topics such as teacher qualifications, professional development experiences, approach to TK, content of instruction, grouping strategies, and strategies for differentiating instruction. Draft surveys will be shared with the funders, TAG members, and select members of the implementation study's stakeholder group, as relevant, for feedback.

Task 5.2: Administer Survey

To minimize burden and to increase efficiency, the TK teacher survey will be administered electronically to the extent possible. We will use a secure, Web-based software program called Vovici, with the use of unique links for each school district to participate. Information about the participant's rights as a research participant will be included at the beginning of the survey, and the respondent will be asked to indicate consent to participate before proceeding. Similar administration procedures will be used to administer the principal and teacher surveys. The AIR team will conduct follow-up, calling and/or e-mailing respondents, as needed, to increase response rates. All follow-up efforts for surveys will be documented in a database and monitored by the project director.

Task 6: District Survey and Follow-Up Interviews

A short district survey will be given to district administrators to track progress on implementation, to be used as context for the study. The survey will be administered in spring 2014 (Cohort 1), and an abbreviated version will be used in spring 2015 (Cohort 2). Additional interviews will be conducted to explore particular topics in more depth, such as how migrant students are served in TK.

Task 6.1: Develop Survey

The district administrator survey will be based on the long-form district survey developed for the implementation study. A subset of items will be taken from the implementation survey to create a shortened version for use in the impact study. Questions will cover topics such as district policies, approach to TK, professional development, and participation in professional learning communities. A draft survey will be provided to the funders, TAG members, and select members of the implementation study's stakeholder group, as relevant, for input, and a final version will be prepared for administration.

Task 6.2: Administer Survey

Like the teacher survey, the district survey will be administered online using Vovici. During the initial recruitment contacts, we will inform the district liaison about the district survey and ask for guidance regarding to whom the survey should be sent. Introducing the survey early in the process will provide time for sufficient follow-up, which is important for district administrators with very busy schedules. We will conduct phone and e-mail follow-up, as needed, and will conduct the short survey over the phone, as necessary, to ensure a high response rate.

Task 6.3: Follow-Up District Interviews on Special Student Populations

District surveys will include questions about the percentage of special student populations not large enough to include in the main outcomes analysis, such as migrant students and students in special education who are served in TK. We also will ask about districts' approaches to meeting the unique learning needs of these students. Districts that have especially high proportions of these students in TK or that have a unique approach to serving them at the TK level will be identified for follow-up phone interviews. Using a semistructured interview protocol, we will explore districts' approaches to migrant education and special education for TK students, their successes and challenges with regard to serving these students in TK, and their perceptions of impacts on these students. This information, combined with exploratory analyses of outcomes for these students in our study sample, if feasible, will provide some descriptive information about migrant students and special education students in TK.

Task 7: Kindergarten Student Assessments

The key source of data for the impact analysis is the one-on-one assessments conducted when the TK children are in kindergarten and with kindergarten students who were not in TK the year prior to kindergarten. Assessments with Cohort 1 students will occur in fall 2014 and in spring 2015; assessments with Cohort 2 students will occur in fall 2015 and spring 2016.

Task 7.1: Schedule Assessments

We will work with a school liaison to schedule the assessments, including identifying convenient times for conducting the assessments to minimize the burden on teachers and the impact on student learning time and finding an appropriate location in which to conduct the assessments. The space should be quiet

(but not private) to minimize distractions for the student yet feel comfortable and safe for the child. An empty classroom, office, resource room, media room, or cafeteria would be appropriate.

On average, approximately 20 students per school (half treatment and half control) will be assessed, and multiple classrooms likely will be involved. We will try to concentrate the assessments over a one-week period, sending multiple assessors to each site as needed. To minimize the variation in assessment timing, we will strive to complete the bulk of assessments in a seven-week window beginning in mid-September. Spring assessments for Cohort 1 will follow approximately six months later, in the March to May time frame. In discussions with the district, we will find out when spring break is scheduled and when the state testing window is for each district and will schedule around these interruptions. We also will explore, with district administrators during our key informant interviews, whether before or after testing would be a preferable time frame for assessments.

Task 7.2: Administer Assessments

Assessments will be administered to students by trained assessors who are part of the research team and are not associated with the school. By a qualified study team member, students will be escorted from their classroom to the assessment space within the school. The assessment space will ideally be a quiet space apart from the classroom, although publicly visible and accessible to all school staff. Assessors will follow strict data collection protocols approved by the IRB, which will include obtaining students' assent to participate and allowing students to terminate the session if they become uncomfortable at any point. Students will be invited to answer some questions and do some activities. They do not have to do this if they do not wish to. Our experience, though, is that students really enjoy doing the assessment activities.

As described previously, we will select a battery of valid and reliable assessments to capture the range of academic and social-emotional outcomes expected to be associated with participation in TK. We will balance an interest in covering multiple constructs with the need to minimize burden on students and schools. Thus, assessment sessions will be limited to 45 minutes, on average. During this time, students will be engaged by developmentally appropriate activities, switching tasks frequently to maintain their attention and interest. Students will be given neutral praise to encourage them without leading their responses in one direction or another. At the completion of the assessment session, students will be given a small thank-you gift, such as a pencil or stickers.

Task 8: Kindergarten Teacher Survey

In addition to the direct assessments of students, we also will ask teachers to rate their students, using one or more teacher report measures. These measures, together with a few questions about teachers' background and classroom characteristics, will be included in a kindergarten teacher survey to be administered in fall 2014 (Cohort 1) and fall 2015 (Cohort 2). In spring 2015 (Cohort 1) and spring 2016 (Cohort 2), previously surveyed teachers (which we expect to constitute the majority of teachers) will get only the student rating items from the survey.

Task 8.1: Develop Survey

The kindergarten teacher survey will include two primary components. The first component will include a set of standard survey items (similar to the TK teacher survey) asking about teacher experiences and qualifications, classroom practices, and approach to teaching kindergarten. The second component will include a set of rating forms that teachers will complete—one form for each student included in the study. The rating forms will focus on social-emotional skills not well addressed by our direct assessments. All items included will be from valid and reliable measures.

Task 8.2: Administer Survey

Like the other surveys described in this study, the kindergarten teacher survey will be administered online. The online surveys will be prepopulated with the students' names, so teachers can simply follow the prompts and rate each student. We will follow up with teachers by e-mail and phone, as necessary, to ensure a high response rate. Teachers also will receive an incentive for completing the surveys.

Task 9: Data Analysis

Task 9.1: Impact Analysis

We will estimate the causal impact of TK participation on children's early language and literacy, mathematics, and social-emotional skills using RD models. Fuzzy RD models will estimate the local average treatment effect, taking into account noncompliance with TK enrollment guidelines. We also will conduct sharp RD analyses that evaluate the impact of offering TK, regardless of whether parents choose to enroll their children. Analyses will examine children's total scores on measures of early language and literacy, mathematics, and social-emotional skills as well as scale scores that examine specific skills in these domains. Analyses conducted in the winter of 2014–15 will examine TK impacts on kindergarten readiness with child assessment data collected in fall 2014 for Cohort 1. In summer 2015, the study team will be able to assess the extent to which TK impacts are sustained at the end of kindergarten, using spring child assessment data collected for Cohort 1 in spring 2015. In the winter of 2015–16, the study team will test the impact of TK for Cohort 2, which will allow us to determine whether the impact of TK increases as the program is fully implemented. We will again assess the extent to which TK impacts for Cohort 2 are sustained at the end of kindergarten, and how these slopes may be different between Cohort 1 and Cohort 2.

Task 9.2: Quality and Implementation Analysis

Using data from the TK classroom observations, the study team will calculate the mean, median, standard deviation, and range of total, and domain/scale scores on classroom quality measures. Surveys of TK teachers will provide an additional source of quality data related to classroom instructional practices and information on structural variables, such as full day versus half day and combination versus stand alone. We will calculate frequencies and means on these structural variables and explore cross-tabulations between process and structural quality measures (e.g., total CLASS scores in full- and half-day classrooms, domain-specific instructional quantity versus quality in combination and stand-alone classroom). Data on TK quality will be collected in spring 2014 for Cohort 1. The study team will clean these data and begin descriptive analyses of these data in summer 2014. Descriptive analyses will be repeated in summer 2015, with TK quality data collected in spring 2015 for Cohort 2.

In the winter of 2014–15, after fall student outcome data are collected for Cohort 1, we will begin to investigate associations between quality and child outcomes, using a series of moderation models as described in the analysis plan. We will conduct similar analyses in the winter of 2015–16, with fall kindergarten data from Cohort 2. Analyses that explore changes in quality from Year 2 to Year 3 of TK program implementation will be our focus in spring 2016. These analyses will describe changes in overall quality and specific domains of quality. We also will explore teacher background characteristics and classroom structure variables that are associated with changes in quality.

Task 9.3: Subgroup Analysis

The study team will run RD models on subgroups, such as English learners, female and male children, children eligible for free or reduced-price lunch, and children who have not attended preschool prior to

school entry. We also may examine subgroups defined by classroom, school, or district characteristics. Examples include children attending TK in combination versus stand-alone classrooms, hub schools versus neighborhood schools, and urban versus rural school districts. To ensure that statistical models have adequate statistical power to detect program effects, subgroup analyses will be conducted with stacked data from both cohorts after Cohort 2 data collection is complete, in spring 2016.

Task 10: Reporting

We will prepare a series of reports and other presentations of findings to share with the funders and other stakeholders, which will include the following:

- Cohort 1 report
- Cohort 1 report update memo
- Final report
- Two research briefs
- Meetings and other presentations, including one published article

Each of these items is described below.

Task 10.1: Cohort 1 Report

Results from the first set of analyses for Cohort 1 will be presented in a Cohort 1 report to be shared with the funders in spring 2015. This report will cover three types of analyses from the first year of data collection. First, we will describe the state of TK in the second year of its implementation. Using the data on classroom quality collected through TK classroom observations and surveys of TK teachers, we will describe the typical TK classroom in terms of instructional quality as well as the range of quality observed across the state. Second, we will address the question of TK impacts on students' readiness for kindergarten, presenting impact estimates based on outcome data for the fall of kindergarten for Cohort 1 students. Third, we will discuss the relationships between quality and outcomes, looking at variation in student outcomes as it relates to the quality of the TK program in which students participated.

Task 10.2: Cohort 1 Report Update Memo

Because the spring assessment data will not be analyzed in time to include in the initial Cohort 1 report, we will prepare a technical memo, providing these results, to follow the Cohort 1 report in the winter of 2015–16. This report will address the question of whether the impacts of TK are sustained through the kindergarten year, by focusing on the impact of participation in TK on the range of outcomes measured in the spring of kindergarten.

Task 10.3: Final Report

We will prepare a comprehensive final report, incorporating all data collection points, which will be submitted to the funders in summer 2016. This report will review the Cohort 1 findings, present Cohort 2 findings, and then present cross-cohort analyses for a complete picture of the quality and impacts of TK. Cohort 2 results will include descriptive analyses of quality in the third year of implementation, impact estimates on fall assessment scores, and an analysis of the relationship between quality and outcomes for Cohort 2. Cross-cohort analyses will bring the two cohorts together to describe: changes in quality over time and any changes in impacts observed over time. In addition, we will describe how impacts of TK

differ for different groups of students, for example, English learners, students receiving free or reduced-price lunch, students without preschool experience, and other subgroups of interest.

Task 10.4: Brief Reports

We anticipate preparing several brief reports that would be accessible to a broader range of stakeholders. First, we will prepare two research briefs, one based on the study results presented in the Cohort 1 report and the other based on the final report. These briefs will be written for a wide audience, including educators, administrators, and policymakers interested in understanding the impacts of TK on student outcomes. We also will prepare two sets of district reports that present results specific to each district participating in the study (assuming numbers of assessed students are large enough). For small districts with too few students to report on, we will provide a small-districts report so that districts have some information that is relevant for their district contexts. By presenting results specific to each district and in comparison to the rest of the sample, we hope these reports will be useful for districts' internal planning and professional development.

Task 10.5: Meetings and Presentations

In addition to formal written reports and briefs, we also will share results with stakeholder groups and other audiences as opportunities arise. For example, these opportunities might include sharing results with the TK Professional Learning Community, the California County Superintendents Educational Services Association, the stakeholder group convened for the implementation study, or at state or national conferences. We also will submit an article for publication in a peer-reviewed journal.

Task 11: Project Management

Project leadership, including Principal Investigator Heather Quick and Project Director Karen Manship, will meet biweekly to ensure that the project is on track in terms of schedule and budget. Regular team meetings also will be held to ensure clear communication across each task team. To ensure quality products, we will participate in AIR's quality review processes, which include periodic project reviews with organizational leadership and reviews of deliverables. We also will meet with the funders by phone, as needed, to discuss progress and any emerging challenges.

During the initial planning phase, we will hire our proposed subcontractor, Survey Research Management (SRM), which will conduct the child assessment data collections.

Exhibit 9 presents the proposed study timeline.

Exhibit 9. Study Timeline

Study Year	Year 1 (2013)				Year 2 (2014)				Year 3 (2015)				Year 4 (2016)			
	Fall	Win	Spr	Sum	Fall	Win	Spr	Sum	Fall	Win	Spr	Sum	Fall	Win	Spr	Sum
Cohort 1 Grade Level	Cohort 1, TK				Cohort 1, Kinder				Cohort 1, 1 st							
Cohort 2 Grade Level	Cohort 2, PK				Cohort 2, TK				Cohort 2, Kinder							
Tasks	Fall	Win	Spr	Sum	Fall	Win	Spr	Sum	Fall	Win	Spr	Sum	Fall	Win	Spr	Sum
Task 1 Planning and Preparation																
1.1 Conduct Key Information Interviews	X															
1.2 Develop Recruitment Materials	X															
1.3 Finalize Measure Selections	X															
1.4 Obtain IRB Approval	X															
1.5 Train Field Staff		X		X		X		X								
Task 2 Sampling																
2.1 Draw District Sample	X															
2.2 Draw School Sample	X	X														
2.3 Identify Pool of Eligible Students			X	X	X		X	X	X							
Task 3 Recruitment																
3.1 Recruit Districts	X	X														
3.2 Recruit Schools		X		X												
3.3 Recruit/Consent Parents			X	X	X		X	X	X							
Task 4 TK Classroom Observations																
4.1 Schedule Observations		X	X				X	X								
4.2 Conduct Observations			X					X								
Task 5 TK Teacher Survey																
5.1 Develop Survey	X															
5.2 Administer Survey			X					X								
Task 6 District Survey and Follow-Up Interviews																
6.1 Develop Survey		X														
6.2 Administer Survey		X						X								
6.3 Follow-Up Interviews on Special Issues			X													
Task 7 Kindergarten Student Assessment																
7.1 Schedule Assessments				X	X	X	X	X	X	X	X	X				
7.2 Administer Assessments					X		X		X		X					
Task 8: Kindergarten Teacher Survey																
8.1 Develop Survey	X															
8.2 Administer Survey					X		X		X		X					
Task 9 Data Analysis																
9.1 Impact Analysis							X		X		X					
9.2 Quality and Implementation Analysis				X		X		X		X	X	X				
9.3 Subgroup Analysis												X	X			
Task 10 Reporting																
10.1 Cohort 1 Report								X								
10.2 Cohort 1 Report Update Memo											X					
10.3 Final Report																X
10.4 Brief Reports									X							X
10.5 Meetings and Presentations	X								X							X
Task 11 Project Management	<i>Throughout Duration of the Study</i>															

Staff and Organizational Experience

Relevant organizational and staff experience are available upon request.

AIR's Institutional Review Board

AIR's institutional review board (IRB) (IRB00000436) has conducted expedited and full-board reviews of research involving human subjects for more than 21 years. AIR is registered with the Office of Human Research Protection as a research institution (IORG0000260) and conducts research under its own Federalwide Assurance (FWA00003952).

AIR's IRB has the capacity to help staff, collaborators, and external organizations comply with federal human subject protection regulations by providing timely reviews. AIR conducts IRB reviews for all of its own internal research activities and provides this service for a variety of subcontractors and collaborators.

All data collection activities in this study will be thoroughly reviewed by AIR's IRB.

Budget

The total budget for this study is **\$7,377,092**.

Totals for each task, by year, are presented in Exhibit 10.

Exhibit 10. Study Budget by Task and Year.

Task	Year 1 (2013)	Year 2 (2014)	Year 3 (2015)	Year 4 (2016)
Task 1- Planning, Preparation and Training	\$79,189	\$325,899	\$513,088	\$222,359
Task 2- Sampling and Collecting Student Record Data	\$24,921	\$111,879	\$103,963	--
Task 3- Recruitment	\$34,149	\$317,959	\$225,833	\$15,517
Task 4- TK Classroom Observations	--	\$347,959	\$428,525	--
Task 5- TK Teacher Survey	\$11,122	\$41,145	\$33,777	--
Task 6- District Survey and Interviews	\$11,122	\$44,493	\$25,433	--
Task 7- Kindergarten Student Assessments	--	\$743,645	\$1,482,545	\$805,262
Task 8- Kindergarten Teacher Survey	\$26,306	\$84,918	\$89,186	\$57,785
Task 9- Data Analysis	--	\$32,845	\$158,031	\$228,154
Task 10- Reporting	\$4,900	\$17,973	\$242,176	\$264,519
Task 11- Project Management	\$40,378	\$66,992	\$62,742	\$50,402
TOTAL	\$232,087	\$2,135,707	\$3,365,299	\$1,643,998

References

- Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability to kindergarten. *Child Development, 78*(2), 647–663.
- Caprara, G. V., Barbaranelli, C., Pastorelli, C., Bandura, A., & Zimbardo, P. G. (2000). Prosocial foundations of children's academic achievement. *Psychological science, 11*(4), 302-306.
- Gormley, W. T., & Gayer, T. (2005). Promoting school readiness in Oklahoma: An evaluation of Tulsa's pre-k program. *The Journal of Human Resources, 40*(1), 533–558.
- Hahn, J., Todd, P., & Van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica, 69*(1), 201–209.
- Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics, 142*, 615–635.
- Jacob, R., Zhu, P., Somers, M.-A., & Bloom, H. (2012). *A practical guide to regression discontinuity*. New York: MDRC. Retrieved from: http://www.mdrc.org/sites/default/files/regression_discontinuity_full.pdf
- Ladd, G. W., & Proffitt, S. M. (1996). The Child Behavior Scale: A teacher-report measure of young children's aggressive, withdrawn, and prosocial behaviors. *Developmental psychology, 32*(6), 1008-1024.
- McClelland, M. M., Cameron, C. E., Connor, C. M., Farris, C. L., Jewkes, A. M., & Morrison, F. J. (2007). Links between behavioral regulation and preschoolers' literacy, vocabulary, and math skills. *Developmental Psychology, 43*(4), 947–959.
- McEwan, P. J., & Shapiro, J. S. (2008). The benefits of delayed primary school enrollment discontinuity estimates using exact birth dates. *The Journal of Human Resources, 43*(1), 1–29.
- Ponitz, C. C., McClelland, M. M., Matthews, J. S., & Morrison, F. J. (2009). A structured observation of behavioral self-regulation and its contribution to kindergarten outcomes. *Developmental Psychology, 45*(3), 605–619.
- Schochet, P. Z. (2008). *Technical methods report: Statistical power for regression discontinuity designs in education evaluations*. (NCEE 2008-4026). Washington, DC: National Center for Education Evaluation, Institute of Education Sciences.



AIR[®]

AMERICAN INSTITUTES FOR RESEARCH[®]

1000 Thomas Jefferson Street NW
Washington, DC 20007-3835
202.403.5000 | TTY: 877.334.3499

www.air.org

Making Research Relevant

LOCATIONS

Domestic

Washington, DC

Atlanta, GA

Baltimore, MD

Chapel Hill, NC

Chicago, IL

Columbus, OH

Frederick, MD

Honolulu, HI

Indianapolis, IN

Naperville, IL

New York, NY

Portland, OR

Sacramento, CA

San Mateo, CA

Silver Spring, MD

Waltham, MA

INTERNATIONAL

Cote d'Ivoire

Egypt

Honduras

Kenya

Liberia

Malawi

Pakistan

South Africa

Zambia