Evaluation of Child Outcomes in Nine Child-Parent Centers: Report for 2015–16

April 2017

Prepared for:
IFF Pay For Success I, LLC
333 S. Wabash Avenue, Suite 2800
Chicago, Illinois 60604
Attention: Matthew J. Roth, Chief Operating Officer
E-mail: mroth@iff.org

SRI Education™
A DIVISION OF SRI INTERNATIONAL

Copy to:
DLA Piper LLP (US)
203 N. LaSalle Street, Suite 1900
Chicago, Illinois 60601
Attention: Richard F. Klawiter, Esq.
E-mail: Richard.klawiter@dlapiper.com

Prepared by:

SRI International
Erika Gaylor
Kate Ferguson
Mary McCracken
Xin Wei
Donna Spiker

Revised June 2019 to correct an error on p. 9

Suggested citation:

Gaylor, E., Ferguson, K., McCracken, M., Wei, X., & Spiker, D. (2016). *Evaluation of child outcomes in nine Child-Parent Centers: Report for 2015-16.* Prepared for IFF Pay for Success I, LLC. Menlo Park, CA: SRI International.

Contents

List of Exhibits

Background

The Child-Parent Center (CPC) model, one of the longest running early childhood intervention models in the United States, has produced some of the most robust long-term academic and social outcomes for children (Reynolds, 2000; Reynolds & Temple, 2008). Beginning in January 2012, as part of a U.S. Department of Education Investing in Innovation (i3) grant to the University of Minnesota, the city of Chicago and Chicago Public Schools (CPS) received funding to (1) increase the number of children who could attend existing CPC sites and (2) increase the availability of CPC programs by adding 16 new sites.

Pay for Success (PFS), previously referred to in this evaluation as a Social Impact Bond (SIB), is a funding mechanism whereby private business and philanthropic partners purchase Social Impact Bonds (SIBs) to support public programs. Government entities, such as a state's Department of Education, pay investors only when a program meets its pre-determined outcomes. The investors bear the full risk of the investment and if a program fails to meet its goals, taxpayers owe nothing. To this end, PFS initiatives typically have an independent evaluator to help determine whether the outcomes have been realized and outcome payments to the private investor need to be made. Beginning in 2014–15, the IFF Pay for Success project funded additional CPC preschool slots at six CPS schools. In 2015–16, three sites (identified by CPS and approved by the city of Chicago) were added to the PFS project. SRI International (SRI) was contracted to conduct the evaluation of the child outcomes for this project, referred to as the "PFS-CPC project." The project is expected to serve four cohorts of preschool children across the nine sites over four school years—Cohort 1: 2014–15, Cohort 2: 2015–16, Cohort 3: 2016–17, and Cohort 4: 2017–18.

This second SRI project evaluation report describes the kindergarten special education placement outcomes of the Cohort 1 children and the kindergarten readiness outcomes for Cohort 2. The report begins with a description of the CPC program and its expansion efforts using PFS funding, including evidence about the impacts of the CPC program model on children's school readiness and school achievement. Next, we describe the approach to the PFS-CPC program evaluation. The last section presents the results of the evaluation for Cohorts 1 and 2.

CPC Program Model

CPC Model Description

The CPC program model is designed to promote school readiness, parent involvement, and early learning that, in turn, will translate into long-term academic achievement, higher graduation rates, and career success for CPC students. The CPC model is unique in that it is designed to (1) provide full- or part-time high-quality preschool experiences for 3- and 4-year-old children and (2) combine those educational experiences with family support services and parent engagement activities. CPC programs deliver synergized services for children and their families from preschool through third grade. Indeed, the CPCs emphasize the provision of comprehensive services and parental involvement—program features that are considered to be strongly associated with program quality (Reynolds & Hayakawa, 2011; Reynolds, Magnuson, & Ou, 2010). A typical CPC site has the components listed in Exhibit 1.

The CPC program model components are explained fully at https://humancapitalrc.org/midwest-cpc/cpc-resources (Human Capital Research Collaborative, 2015). For this report, the components listed in Exhibit 1 were taken from the draft evaluation plan in the PFS-CPC expansion agreement (see Appendix A, pp. 9–11). Note that the CPC model as conceptualized in the current PFS expansion project focuses primarily on providing high-quality preschool education, engaging parents in their child's education through a parent resource teacher (PRT) at the child's preschool, and promoting continuity and stability from pre-K through the primary grades. Because the focus for the PFS-CPC project is on preschool programming, SRI's evaluation is designed to measure the impact of the preschool components on children's short- and long-term outcomes.

Exhibit 1. CPC Program Model Components

Effective Learning Experiences

- Offer pre-K classes that are limited to 34 children for half-day classrooms (two sessions of 17 children each) and have a minimum of two teaching staff. Full-day classrooms, if available, will be limited to 20 children per session.
- Provide highly qualified educational staff who will deliver the classroom instruction and parent engagement activities. For example, classroom teachers are certified with a bachelor's degree (or higher). Overall, program staff must adhere to the requirements set forth by the CPS Talent office, in accordance with collective bargaining unit agreements, and state regulations. Any changes in CPS education and certification requirements will be complied with.
- Use data to drive instruction by effectively documenting the organization and implementation of instructional practices to monitor quality and adherence to the program, which is completed by all program staff where appropriate.
- Program staff meet with parents over the course of each school year to review their child's progress and discuss parent program opportunities with the Parent Resource Teacher (PRT).

Aligned Curriculum

- Implement a CPS District curriculum and formative assessment that are aligned to standards, domains of learning, assessments, and learning activities.
- Collaborate with the PRT and classroom teachers to ensure that opportunities to engage families in student learning are available, appropriate, and aligned to the program and parents' needs.
- CPS and, most specifically, the Office of Early Childhood Education provide meaningful professional development and ongoing coaching and feedback for teachers, aides, and other staff members that facilitates high-quality instructional practices.

Parent Involvement and Engagement

- Engage a PRT and School-Community Representative to work closely with the head teacher and liaisons to maintain a consistently supportive parent program.
- Encourage parents to sign a CPC school-home agreement at the start of the school year outlining a plan for fostering learning at home and participating in CPC activities.
- Offer and engage families in monthly activities. PRTs create and distribute a monthly parent involvement calendar and conduct parent/teacher conferences over the year to review progress in the parent program.
- Provide a resource room dedicated to parent and family activities through kindergarten when possible.
- Provide culturally responsive learning opportunities for families that provide flexibility for families' needs and schedules.

Collaborative Leadership Team

- Engage a program leadership team that includes the head teacher, PRT, and school-community representative.
- Meet regularly, under the direction of the principal, to discuss operations and best practices within the CPC.
- Meet regularly, under the direction of the Office of Early Childhood Education (OECE) management team, with staff from across sites to share challenges, experiences, and best practices and make frequent on-site visits to monitor the quality and effectiveness of the program.
- Establish meaningful partnerships with community providers to strengthen service delivery and enlist local universities in training opportunities.

Exhibit 1. CPC Program Model Components (concluded)

Continuity and Stability

- CPC pre-K classrooms are collocated in the same building as kindergarten classrooms, when possible, to promote familiarity and integration for students as they transition to kindergarten.
- Provide a structure of communication, planning, and joint activities under the direction of the principal, leadership team, and OECE management team from pre-K through the primary grades.
- Provide a part-time kindergarten aide when funding is available to support the transition into kindergarten.

Professional Development System

- Offer ongoing professional development opportunities on current trends and needs in early childhood education classrooms, through the OECE and the CPC leadership teams, including topics such as quality curriculum and instruction, data-driven instruction, learning environment, social and emotional needs, and parent engagement.
- Meet regularly and create professional learning communities to review ways to support instruction in the classroom and with other teachers.

Source: Adapted from Chicago Child-Parent Center Social Impact Bond Evaluation Plan, dated December 2, 2014 (in Appendix A).

Expected Outcomes from the CPC Program Model

SCHOOL READINESS

Previous research on the CPC program showed significant positive effects on children's kindergarten readiness (Reynolds, 1995; Reynolds, Temple, Robertson, & Mann, 2002). Examination of a more recent cohort of CPC participants indicated that they had significantly higher scores on a measure of language proficiency at the end of the program than children enrolled in other publicly funded preschool programs (Reynolds, 2002). More recently, Reynolds and colleagues reported that CPC participants are more likely to meet kindergarten readiness standards in four of six educational focus areas, or 'domains' on a teacher-rated measure (70%) compared with preschool children in the school district who did not attend CPC preschool classrooms (52%) (Reynolds, Richardson, Hayakawa, Englund, & Ou, 2016).

THIRD-GRADE READING AND LITERACY

The Chicago Longitudinal Study (CLS) followed children over time using administrative records to examine attendance, achievement, and graduation rates in CPC participants compared with children who did not attend CPC preschool. One study found a significant positive impact on third-grade reading achievement for pre-K to third-grade participants (.53 standard deviation) compared with participants who attended CPC only for pre-K and kindergarten (Reynolds, 1994). Smaller studies of

high-quality preschool interventions have found similar impacts on later school achievement compared with a no-preschool Comparison group (e.g., the Abecedarian study: Campbell, Ramey, Pungello, Sparling, & Miller-Johnson, 2002; Perry preschool project: Belfield, Nores, Barnett, & Schweinhart, 2006).

REDUCED SPECIAL EDUCATION USE

The long-term CLS study showed that extended CPC participation (defined as 4 to 6 years) resulted in reductions in the use of special education. Among children 6 to 18 years old, CPC participants had an average rate of special education placement of 14.4% compared with 24.6% for children in the comparison group (who did not attend CPC preschool), indicating that CPC participants had a 41% lower rate of special education placement (Reynolds, Temple, & Ou, 2003). This finding is consistent with another analysis using the CLS sample that compared the average rates of special education placement over time for children who had attended a CPC preschool with those of children who attended a full-day non-CPC kindergarten classroom: special education placement rates of 12.5% and 18.4%, respectively (Conyers, Reynolds, & Ou, 2003). It is noteworthy that these estimates average special education placement rates over a wide age range extending beyond the early school years. A more recent study of North Carolina's state-funded preschool program used statewide population data from 1995 to 2010 to show that third-grade special education rates were reduced by as much as 39% for children who participated in the preschool program, even after taking into account a variety of child and family risk factors, types of special education categories, and funding levels that varied by year (Muschkin, Ladd, & Dodge, 2015). Other reviews of a variety of preschool program models also reported reductions in special education placement as one of the many cost savings results from participation in high-quality preschool programs like the CPC model (Karoly et al., 1998; Lynch, 2007).

In summary, positive impacts on kindergarten readiness, third-grade reading achievement, and special education placements have been cited extensively to demonstrate the short- and long-term benefits for the individual child and savings for society that come from investing in early childhood education. These studies were used as the basis for identifying the selected outcomes in the current evaluation and for calculating the repayments that will be made in the Chicago PFS-CPC project.

Chicago PFS Project (PFS-CPC Project)

During 2015–16, the PFS expansion of the CPC model involved funding for part-day or full-day CPC preschool at nine sites. Exhibit 2 indicates the year each site began receiving PFS funding and whether the site expanded an existing CPC program or began implementing the CPC program for the first time (i.e., a "new" CPC site).[1] In 2015–16, the PFS funding provided preschool to an additional 782 3- and 4-year-olds across the nine sites (see Exhibit 3). The funding paid for the expansion of classroom programming at each site as well as enhanced resources and instructional materials to implement the CPC model. The CPC program typically serves both 3- and 4-year-olds, serves both 3- and 4-year- olds, sometimes in mixed-age classrooms. Thus, the investor funding was used to provide CPC preschool and enhanced services to both 3- and 4-year-olds.

Exhibit 2. Description of Participating CPC Sites, by Project Year

Site	Year 1 2014–15	Year 2 2015–16
De Diego	Expanded	Continued
Peck	Expanded	Continued
Melody	Expanded	Continued
Fiske	Expanded	Continued
Thomas	Expanded	Continued
Hanson Park	New[2]	Continued
Edwards		Expanded
Tonti		New
Davis		New

Note: "Expanded" indicates that a site used PFS funding in that year to expand an existing CPC program, "New" indicates a program received PFS funding to begin implementing a CPC program for the first time, and "Continued" indicates that a site continued to receive PFS funding for an additional year.

[1] Three of the six sites in Year 1 had been providing CPC services since 2012, at the start of the i3 federal grant, and two had been providing CPC services since 2013, when the original sites from the 1970s were merged with the current site. For the three additional sites in year 2, two were new to providing CPC services and one had been providing CPC services since 2012.
[2] This site only operated for half of the year due to delays in hiring. Thus, the first full year of implementing CPC was 2015–16.

The project administrators anticipated that four cohorts of children will be served across the nine sites, identified by the school year in which children begin preschool (Cohort 1: 2014–15, Cohort 2: 2015–16, Cohort 3: 2016–17, Cohort 4: 2017-18) (see Appendix B for grade levels of children in the four cohorts across years.)

Evaluation Design

Because government pays investors only when outcomes are achieved, PFS initiatives typically have an independent evaluator to help determine whether the outcomes have been realized. SRI is conducting the independent evaluation of the outcomes of the PFS-CPC expansion project. The SRI evaluation team developed the evaluation methodology building on a draft evauation design written by a team that included the Harvard Social Impact Bonds Technical Assistance Lab. The project also has an oversight committee of early education and research experts. The evaluation team has been charged with independently documenting the outcomes-based performance measures of the initiative. This kind of evaluation is not intended to test the impact of the CPC model against other preschool models; rather it is tracking the outcomes of the participating children against specific performance standards. The evaluation is addressing three performance questions:

1. What is the rate of kindergarten readiness in children participating in the PFS-CPC sites as defined by performance on the Teaching Strategies (TS) *GOLD*™ instrument (completed by teachers in the spring of the preschool before a child enters kindergarten)?
2. What is the rate of third-grade literacy as defined by performance in meeting or exceeding grade-level performance on the state- or district-administered third-grade assessment in reading?
3. What is the rate at which students are identified with special education needs and placed in special education services (starting in kindergarten) compared with a matched comparison group of children?

Kindergarten readiness is being measured in the spring of preschool for CPC participants (as described below), and third-grade literacy will be measured in the spring of third grade after the administration of required state achievement tests. SRI will begin measuring special education placement in kindergarten and repeat its evaluation annually until spring 2020 (in spring 2020, Cohort 1 will have reached the

fourth grade; Cohort 2, the third grade; Cohort 3, the second grade; and Cohort 4, the first grade).[3]

For the evaluation of the PFS-CPC project, SRI is using two designs to track the primary outcomes: a descriptive study for the *kindergarten readiness* and *third-grade literacy* outcomes and a quasi-experimental design for the *special education outcomes from first to fourth grades* (see analysis approach for further information). Specifically, for the kindergarten readiness and third-grade literacy outcomes, there will be no comparison group for evaluating the outcomes and calculating the subsequent repayment. Evaluation of these two primary outcomes will be based on the intervention group only, and payments will be calculated using outcomes relative to national standards. In the planning phase, it was determined that both the kindergarten readiness and literacy outcomes had normative information so that children's performance on these measures could be used to identify whether they were performing at or above normative trends for comparable age their peers. The decision was to use this kind of standard rather than compare performance with that of a comparison group of children. In addition, the kindergarten readiness data are not available for children with no preschool experience, given that the kindergarten readiness measure is collected during the spring of pre-K in Chicago Public Schools and only for children who attended preschool.

For special education outcomes (first to fourth grades), children are identified as receiving the intervention (i.e., attending a CPC preschool classroom) in the year they are in preschool and then are matched to children with similar demographic characteristics but who did not attend any preschool (CPC or otherwise) in CPS. This no pre-K comparison group is identified when the children are in kindergarten for each of the four cohorts. That is, SRI will create a no pre-K comparison group for each cohort of intervention children using propensity score matching processes.

Analysis Approach

CPC INTERVENTION SAMPLE INCLUDED IN ANALYSIS

Children were included in the intervention cohorts if they attended one of the PFS-CPC sites that was fully implementing the CPC program during their preschool year,[4]

[3] Further evaluation will be required after 2020.
[4] There were six sites for Cohort 1 and nine sites for Cohort 2.

were enrolled in either a full- or half-day pre-K classroom, were not identified as having a severe disability, were income eligible (i.e., eligible to receive free or reduced-price lunch), and were at least 4 years old in September of their preschool year. Additionally, a child needed to have attended a CPC pre-K classroom for at least 66% of the days (not consecutively) in a given school year—a percentage considered a sufficient amount or dose of the intervention to affect child outcomes.

The project is based on the hypothesis that high-quality early childhood education will prevent or reduce a future need for special education services for children considered at-risk for developing delays or mild disabiltiies. As such, children diagnosed with severe disabilities were excluded from the project. Early childhood education and intervention also may reduce the need for children with mild delays or speech and language impairments in preschool from needing additional special education services in kindergarten and beyond. The project does not expect to prevent children with severe disabilities or needs from receiving special education services. Children were categorized as having no disability, a mild disability, or severe disability based on *a priori* decisions of the evaluation team in the planning and evaluation design phase. A severe disability could include autism, specific learning disability, deaf-blindness, deafness, hearing impairment, orthopedic impairment, other health impairment, traumatic brain injury, visual impairment, and multiple disabilities. A mild disability could include developmental delay, speech and language impairment, specific learning disability, and educational support accommodations or modifications for children with no other disability (mild or severe).[5] Additionally, children were excluded from the intervention cohort if they were in a separate classroom for students with special education needs. Finally, children were excluded if they were identified with special needs and already had an Individualized Education Plan (IEP) prior to starting their preschool year at age 4 and were specifically assigned to one of the CPC sites because the site had blended classrooms (i.e., based on a CPS district policy, some school sites had general education classrooms with additional supports to better serve children wth IEPs, which are referred to as “blended classrooms”).

Each cohort includes children from all PFS-funded sites that were providing the CPC model to 3- and 4-year-olds during that cohort’s 4-year-old preschool year. Inclusion

[5] In an earlier version of this report, the “mild” disability category was incorrectly described as including SPL, DD, and ED. This description has been corrected; the results of analyses have not changed.

of all eligible 4-year-olds in this group increases the sample size for the study to provide a more reliable and valid assessment of kindergarten readiness at the CPC sites. All the children across all classrooms received the full CPC model. That is, the experience of all 4-year-olds enrolled in these CPCs is similar, with a common curriculum, professional development, and parent engagement aligned through monthly Collaborative Leadership Training by all CPCs, including high-quality preschool and family support services and parent engagement activities. Thus, the evaluation does not distinguish between preschool slots funded by PFS versus other CPC funding sources.

Each PFS-CPC cohort is defined as meeting the eligibility criteria above and will become the cohort to be tracked for outcomes in kindergarten and in later grades. Each cohort also will be used to identify a matched comparison group of children in kindergarten for comparing special education outcomesat the end of kindergarten and in later grades.

Exhibit 3 shows the enrollment information for CPC preschool slots at PFS-funded sites, by year. At the end of Year 2 (the 2015–16 school year), administrative enrollment data showed that 1,378 3- and 4-year-old children were attending preschool at these nine sites (537 3-year-olds; 841 4-year-olds). PFS expansion funding covered the costs of providing CPC preschool for 782 of these 1,378 children.

Exhibit 3. Enrollment at PFS-funded CPC sites, by year

	Year 1 Enrollment (2014–15)	Year 2 Enrollment (2015–16)
Number of sites	6	9
Total enrollment	**653**	**1,378**
3-year-olds	*267*	*537*
4-year-olds	*386*	*841*
PFS-funded seats	374	782

For Cohort 2 intervention group, SRI requested a data export that included all students ever enrolled as grade PK (the CPS designation for 4-year-olds in preschool) in the PFS-funded CPC sites at any time in the 2015–16 school year.

Overall, 1,004 PK students were ever enrolled at one of the nine CPC sites during 2015–16.[6] Across the total sample of 1,004 PK children, 654 or 65% met all the eligibility criteria. Exhibit 4 indicates the exclusions from the original sample of PK children ever enrolled in one of the sites that resulted in the final sample of children included in the analysis for each cohort.

Exhibit 4. Children Attending PFS-funded CPC Sites, by Cohort and Exclusion Criteria

	Cohort 1 2014–15 (*n* or %)	Cohort 2 2015–16 (*n* or %)
Number of sites included	5	9
Number of four-year old children attending these sites (grade PK)	**449**	**1,004**
Number of four-year old children meeting eligibility criteria	**313**[7]	**654**[8]
Reason for exclusion	%	%
Did not attend 66% of days	21%	21%
Severe disability and/or enrolled in separate special education classroom	3%	2%
Had an IEP prior to PK year	3%	7%
Not eligible for free or reduced-price lunch or insufficient documentation	2%	1%
Under 4 years old in September of PK year	<1%	4%

Cohort 1 (2014–15) included 313 children, and Cohort 2 (2015–16) included 654 children. Meeting the attendance criterion was the greatest challenge, with approximately 80% of Cohort 1 and 79% of Cohort 2 PK children ever enrolled attending for 66% of the days. The demographic characteristics of Cohorts 1 and 2 are described in Exhibit 5. Cohort 1 is also described in the year 1 report (Gaylor et al., 2016). The final samples of children who were included were similar in many ways to the children who did not meet the eligibility criteria with a few exceptions. The

[6] The number of children *ever* enrolled is different from enrollment estimates at any given point in the year. As children left a CPC site, new children were enrolled. The 1,004 includes all children *ever* enrolled during the 2015–16 year. Based on enrollment in May/June 2016, CPS reported that 841 4-year-old children were enrolled at the nine CPC sites at the end of the year.
[7] Of the 313 children selected for Cohort 1 in their PK year (2014-15), 289 were enrolled in CPS on the 20th day of school during their kindergarten year, for a Cohort 1 kindergarten retention rate of 92.33%.
[8] Of the 654 children selected for Cohort 2 in their PK year (2015-16), 619 were enrolled in CPS on the 20th day of school during their kindergarten year, for a Cohort 2 kindergarten retention rate of 94.65%.

children who were included in each intervention cohort were more likely to be Hispanic and more likely to speak Spanish at home compared with the children who were excluded from that cohort.

Exhibit 5. Intervention Sample Characteristics, by Cohort

Characteristic	Cohort 1 2014–15 (*n* = 313) (percent)	Cohort 2 2015–16 (*n* = 654) (percent)
Male	50	48
Hispanic	67	79
African-American	30	17
Caucasian	1	2
Asian or Multiracial	1	<1
Designated as English Language Learner (ELL)	44	57
Identified mild developmental delay or disability	4	4
Enrolled in full-day Pre-K classrooms	37	40

ANALYZING IMPACT ON KINDERGARTEN READINESS

SRI examined kindergarten readiness using Teaching Strategies (TS) *GOLD*™ scores from the spring before the child entered kindergarten.[9] TS *GOLD* is a teacher-reported measure of young children's skills across six developmental domains: literacy, language, mathematics, cognitive development, socio-emotional well-being, and physical health. We are using this measure because it was the only available child assessment that CPS routinely collects and was therefore selected as the measure of kindergarten readiness by the PFS planning team.[10] It is used routinely in the CPS preschool programs, and there is no CPS-wide measure of kindergarten readiness that is completed as children are entering kindergarten in the fall of the school year.

[9] The TS *GOLD* assessment was developed to monitor the skills of children attending a child care or preschool program so teachers can adjust their instructional strategies depending on the children's progress on a variety of skills and behaviors. TS *GOLD*™ was not developed as a measure of kindergarten readiness.

[10] The methodology involved in SIB projects relies on use of available administrative data rather than additional data collection to evaluate outcomes.

Calculating Impact on Kindergarten Readiness

As described, we calculated the impact of CPC on kindergarten readiness by comparing children's performance on the measure with national norms. We selected this approach for two reasons. First, adequate normative data enables us to identify whether children in the sample were performing at or above a widely accepted standard. Second, creating an appropriate comparison group within CPS was not possible; kindergarten readiness data are not available for children without preschool experience (our comparison group), given that the kindergarten readiness measure is collected during the spring of pre-K in Chicago Public Schools.

The metric for kindergarten readiness is the percentage of children who are performing at or above national trends across at least five of these six domains.[11] A child is determined to be ready for kindergarten if he or she is rated by the teacher as demonstrating levels of skill or knowledge that are expected for a child at a particular age; the reference point for such expectations come from the observed abilities of other children from a representative sample of same-aged peers in the United States. We categorized children as kindergarten ready on each domain by the criterion of meeting or exceeding the 50% percentile on the standard score for that domain using scores from the most recently published technical manual (Lambert, Kim, & Burts, 2014a). Then, we calculated the percentage of children who met this criterion on five of six domains.[12]

ANALYZING IMPACT ON SPECIAL EDUCATION PLACEMENT

Special education placement was determined using data on children's disability designation at any time during the child's kindergarten year. Children were classified into three categories: children receiving special education services for mild disability,

---

[11] No data are available on which domains of the TS *GOLD* assessment to use to reliably and validly determine kindergarten readiness. The decision to define kindergarten readiness as performing at or above national trends on five of six domains (and not four of six) aligns with the National Research Council's definition of school readiness, which includes age-level skills across multiple domains (National Research Council, 2008). The threshold of five of six domains also takes into account that a child may not meet a standard for all six domains, especially in the spring of preschool, as these skills are emerging during this time period.

[12] Teacher-reported assessments have some unknown sources of variability, and the GOLD assessment is no different. Research on the GOLD assessment indicates that between 17% and 25% of the variance in scale scores is accounted for by unmeasured differences between classroom and teachers, including rater effects (Lambert et al., 2014). At this time, there is no consensus on how to calculate kindergarten readiness using GOLD assessment scores. Thus, we continue to use the a priori definition and benchmark.

children receiving special education services for severe disability, and children not receiving special education services (no IEP). Recall that our hypothesis is that high-quality preschool via the CPC program will decrease the chances that children who are at risk will need special education services in the future. Because we are not trying to prevent children with severe disabilities from receiving the special education services they need, we restricted our definition of special education outcomes to children who needed special education for mild delays or disabilities defined as those children who had an IEP for the following: speech and language issue (S/L), developmental delay (DD), emotional disturbance (ED),[13] which is the only information available in the administrative dataset describing the type and severity of disability. This helps avoid the perverse incentive of withholding special education services from children with severe disabilities. We report the special education placement outcomes during kindergarten for children who attended CPC preschool in 2014–15 versus their peers who did not attend any preschool administered by CPS.

The effect size of the impact on special education placement for Cohort 1 was calculated using the risk difference approach. The equation is the following:

$$ES_{i,t} = SPED_{C,i,t} - SPED_{T,i,t}$$

where $ES_{i,t}$ is the effect size for cohort *i* in year *t*, $SPED_{C,i,t}$ is equal to the average of a binary indicator of special education placement among the no pre-K comparison group for cohort *i* in year *t,* and $SPED_{T,i,t}$ is the average of a binary indicator of special education placement among the intervention group for cohort *i* in year *t.* The same calculation will be used for each cohort for each year through sixth grade as described below.

Based on conversations with special education experts and reviewing existing CPS data, the consensus by the planning committee is that the vast majority of children who have a disability will be identified by the end of sixth grade (Blackorby et al., 2010). As a result, after the sixth-grade effect size has been calculated, IFF (or the district) will average the effect size over the last 3 years (fourth, fifth, and sixth grades) and lock in that average effect size for the purposes of calculating payments in grades 7 through 12. This lock-in rate will be calculated separately for each

[13] Note that no children were identified as needing special education services for an Emotional Disability or Distrubance (ED) in the intervention cohort or the comparison group.

intervention cohort. SRI may propose changes to this lock-in methodology in the event that the team determines that it produces skewed results. Any modifications must be approved by CPS, the city of Chicago, the project coordinator, and approved by the lender committee.

SELECTING A MATCHED COMPARISON GROUP

For the special education placement outcome, we conducted propensity score analysis to identify an appropriate comparison group that had not received CPS preschool in either school- or community-based settings. Propensity score methods are quasi-experimental approaches that were developed to approximate findings obtained from randomized controlled trials (Becker & Inchino, 2002). They have been used increasingly in analysis of observational data to reduce selection bias in estimating treatment, policy, or intervention effects when randomized controlled trials are not feasible or ethical (Rosenbaum & Rubin, 1983, 1984, 1985). In essence, propensity score methods help to identify a comparison group that mimics what might have been obtained using random assignment.

To create the comparison group, we first restricted our data set to all kindergarteners in CPS who were 5 years old or older on September 30, were eligible for free and reduced-price lunch, and did not attend preschool in CPS (either in a CPC or other CPS preschool classroom),[14] and who were not attending kindergarten at a school with a CPC program.[15] This reduced our potential sample to approximately 10,562 or about one-third of the total number of children enrolled in kindergarten in CPS in 2015–16 (approximately 30,000). The sample was further reduced to 9,445 who had no missing data on any baseline covariates and outcome. We then applied propensity score analysis to identify a propensity score for each of the 9,445 children eligible to be in the comparison group. The propensity score is the predicted probability of being in an intervention based on a set of potentially confounding covariates (e.g., child and

[14] The evaluation team only had data about whether children had received preschool in CPC or other CPS-sponsored classrooms. As such, it is possible that some children in this comparison group may have participated in a preschool program such as Head Start outside of the district in a community-based setting.

[15] The initial planning team had suggested to also exclude children who were enrolled in charter schools, magnet or selective enrollment schools, and schools that serve exclusively a special education population. However, there are no elementary schools that serve exclusively a special education population. The evaluation team did not think it necessary to exclude children who attended charter or magnet schools because we did not have adequate information showing these were better schools than the “business as usual” elementary schools children could have attended.

neighborhood background characteristics; see below for more detail) using logistic regression. The key advantage of using a propensity score is the ability to balance intervention and comparison groups on a large number of covariates by using a linear combination of covariates for a single score. Simply, the propensity score is a measure of how similar children from the comparison group are to the children in the intervention group on a large number of covariates.

We applied the propensity score analysis in two ways: propensity score matching (PSM), and propensity score weighting (PSW). First, we created a propensity score match which identified the closest single match for each of the Cohort 1 Intervention children. This one-to-one match identified 297 good matches for 297[16] Cohort 1 Intervention children.[17] However, PSM has a drawback of not being able to use all the subjects in the comparison pool; thus reducing statistical power (i.e., the evaluation's ability to detect meaningful differences between groups). To increase power and include as many cases in the comparison and intervention groups as possible, we then applied an alternative method: propensity score weighting (PSW). PSW has the advantage of including all eligible children in the comparison group sample rather than only matched cases. It weights each comparison child by their propensity score, a measure of similarity between intervention and comparison on a large number of covariates; comparison children were weighted higher if they were more similar to intervention children and were weighted lower if they were to less similar to Intervention children.

The way PSW works is that each child is given a weight derived from logistic regression which represents how closely the child matches the intervention group; in this case, how well-matched they were on the selected child and neighborhood

[16] Although the intervention Cohort 1 started with 328 children, 15 children were removed because they had an identified IEP prior to starting grade PK and were assigned to these sites prior to the intervention. This issue was not identified until after the Year 1 report was finalized. Then, 16 children from the remaining 313 were no longer enrolled in the district in 2015-16 leaving a final Cohort 1 sample for kindergarten of 297.

[17] We did not use PSM to identify a larger number of children (e.g., 1,200) as described in the evaluation plan because often PSM alone does not produce an equivalent sample of equally good matches. For example, we could try selecting four comparison children for every intervention child, but 1 or 2 of those children could be really close matches and the other 2 could be less close matches. In the end, it's harder to get matched groups using PSM which is why we decided to use PSW approach. The PSW is a better approach to identifying a well-matched group for comparing outcomes. We also decided to use a larger comparison sample and propensity score weighting in order to mitigate the impact of children in the comparison group who may have received pre-K through Head Start or other programs outside CPS. Results were similar across propensity score matching and weighting approaches.

characteristics. The weight is not related to the outcome (special education status). This propensity score weighting approach adjusts for confounds using inverse propensity score estimators, as recommended by Curtis, Hammill, Eisenstein, Kramer, and Anstrom (2007), Hirano, Imbens, and Ridder (2003), and Rosenbaum and Rubin (1983). From the logistic regression model, we calculated a probability that the comparison child would be in the intervention group. The weight for intervention students was 1.0, and the weight for comparison students was equal to their propensity score transformed to an odds scale (pi/1-pi) (Harder, Stuart, & Anthony, 2010; Hirano et al., 2003).

In the weighting approach, children in the comparison group who are more like the intervention group children are weighted more heavily, and comparison group children less like the intervention group children get smaller weights. Outcome data for each child is given a numerical weight based on their baseline demographic characteristics. For example, if children in the intervention group are more likely to be Hispanic than White, the data for a child who is White will be weighted less in the analysis than a child who is Hispanic.[18] The final comparison sample comprises all 9,445 children, weighted to closely match children in the intervention group.

An important aspect of estimating the propensity score is the selection of covariates. Researchers suggest that covariates that affect both intervention participation and outcomes should be included in the estimation of the propensity score (Caliendo & Kopeinig, 2008; Heckman, Ichimura, Smith, & Todd, 1998; Lechner, 2002; Ravallion, 2001). Covariates included in this study were selected based on findings from studies that have examined neighborhood effects associated with child outcomes (Harding, 2003; Root & Humphrey, 2014; Sampson, Sharkey, & Raudenbush, 2008; ten Bensel, Gibbs, & Lytle, 2015). Our covariates came from four data sources: school district data for the 2015-2016 school year, census tract data for 2013, community area public health data for 2009, and police district crime report data for 2010. All covariates used the most recently available data at the time of analysis. The PSW did result in well-match intervention and comparison groups. Additional details about the PSW analysis are contained in Appendix C.

[18] This is an oversimplification because the approach actually uses all of the demographic data available to create the weight for each comparison child. That is, instead of matching on each of the covariates; children are given a weight that is based on the combination of all of the covariates.

Results

We first present the kindergarten readiness results for Cohort 1 and Cohort 2, and then the results for kindergarten special education placement outcomes for Cohort 1 children.

Kindergarten Readiness

COHORT 1 KINDERGARTEN READINESS

The TS *GOLD*™ Spring 2015 data were not available for three[19] of the 313 children in Cohort 1, resulting in a final *analytic* sample for this outcome of 310 children (99% of the 313 children), which we used to calculate kindergarten readiness. Of those 310 children, 61% (60.97%)[20] were considered to be ready for kindergarten, where "readiness" was defined as scoring at or above the 50th percentile on at least five of the following six domains: literacy, language, mathematics, cognitive development, socio-emotional well-being, and physical health. One-tenth (9%) of the 313 children did not score at or above the 50th percentile for any domain, with 3% meeting the criteria for only one domain, 7% for two domains, 12% for three domains, and 9% for four domains (Exhibits 6 and 7). Additionally, children who attended full-day CPC preschool had higher rates of kindergarten readiness (66%) compared to children who attended half-day CPC preschool (58%).

COHORT 2 KINDERGARTEN READINESS

The TS *GOLD*™ Spring 2016 data were not available for 68[21] of the 654 children in Cohort 2, resulting in a final *analytic* sample for this outcome of 586 children (90% of the 654 children), which we used to calculate kindergarten readiness. Of those 586 Cohort 2 children, 42% (41.64%) were considered to be ready for kindergarten, where "readiness" was defined as scoring at or above the 50th percentile on at least five of the following six domains: literacy, language, mathematics, cognitive

---

[19] These children were missing data either because they were no longer enrolled in the spring (*n* = 2) or their GOLD assessment was incomplete (*n* = 1).

[20] In the year 2 report, we revised who was included in the intervention cohort as we learned in 2016-17 about the children identified prior to PK who were assigned to the Cohort 1 sites because of the blended classrooms.

[21] These children were missing data either because teachers rated their Literacy and Language domains in Spanish, for which scale scores are not available (n = 61), they could not be found in the GOLD™ system (n = 5), or their GOLD™ assessment was incomplete for unknown reasons (*n* = 2). All of the 61 children who were rated only on their Spanish language and literacy attended two of the sites added in year 2.

development, socio-emotional well-being, and physical health. One-seventh (14%) of the 586 children did not score at or above the 50th percentile for any domain, with 11% each meeting the criteria for one, two, three, and four domains (see Exhibits 6 and 7). Additionally, children who attended full-day CPC preschool had higher rates of kindergarten readiness (50%) compared to children who attended half-day CPC preschool (36%).

Exhibit 6. Percent of Children Meeting Kindergarten Readiness Criteria, Across Domains, by Cohort

Number of domains meeting or exceeding the 50th percentile	Cohort 1 2014–15 (percent)	Cohort 2 2015–16 (percent)
0	9%	14%
1	3%	11%
2	7%	11%
3	12%	11%
4	9%	11%
5	10%	19%
6	51%	23%
5 or 6	**61%**	**42%**

Exhibit 7. Percent of Children Meeting Kindergarten Readiness Criteria, by Cohort and Domain

Domain	Cohort 1 2014–15 (percent)	Cohort 2 2015–16 (percent)
Cognitive	82%	65%
Language	66%	50%
Literacy	75%	62%
Math	81%	72%
Physical	60%	30%
Social-emotional	79%	63%

Special Education Placement

COHORT 1 SPECIAL EDUCATION PLACEMENT

After ensuring the two groups were equivalent on child and neighborhood characteristics and weighting appropriately, we examined the special education placement rates for mild and moderate developmental delay or disability for Cohort 1 and its comparison group.[22] The special education rate was slightly lower in the intervention group than the comparison group (4.38% for Cohort 1 in kindergarten and 4.94% for children in the comparison group in kindergarten). This is a difference of 0.56 percent. All of these children were either categorized as having developmental delay or speech and language impairment.

Discussion

Sociodemographic risk factors—the most extensively studied of which is poverty—are highly predictive of developmental trajectories. Children from low-socioeconomic-status (SES) households are less likely to enter kindergarten with the pre-academic and social skills needed to succeed and are more likely to require later special education services later (Isaacs, 2012; Brooks-Gunn, Rouse, & McLanahan, 2007; Lee & Burkham, 2002; Hogan, Msalll, Rogers, & Avery, 1997). Early childhood

[22] All students in the comparison group with an IEP were identified during their kindergarten year, In the intervention group, 3 students were newly identified during their kindergarten year (23%), and 10 continued to have an IEP which was identified during their PK year (77%).

programs potentially mitigate the risks endemic to children from disadvantaged backgrounds, with studies showing that the strongest positive short- and long-term outcomes result from intensive and comprehensive programs targeting low-income children (Burger, 2010; Institute for Research on Poverty, 1997; Reynolds et al., 2010). Indeed, prior studies have highlighted early childhood as a critical and sensitive period for the development of brain architecture and neurochemistry (e.g., Knudsen, Heckman, Cameron, & Shonkoff, 2006) and subsequent academic and socio-emotional well-being (Shonkoff & Phillips, 2000).

First implemented in Chicago in 1967, the CPC model has a long history of offering innovative, targeted approaches to school reform including a comprehensive system of educational and family support services during the preschool through third grade years for young children in low-income neighborhoods. The intervention promoted young children's school success through language enrichment and intensive, mandatory parent involvement within a system of comprehensive support services for children and their families. The CPC model, integrated into the CPS system since its inception in 1967, has been systematically evaluated for its impact on child and family outcomes. A notable by-product of the CPC program's efforts is the Chicago Longitudinal Study (CLS), which has supported researchers' efforts to develop a deeper understanding of the "active" ingredients of early dual-generation interventions and early childhood interventions more generally. The following are key relevant findings from analyses conducted on the CLS samples:

- Nearly half of children (44%) attending a CPC for one year were considered ready for kindergarten compared with 28% of children who had no preschool (unpublished data, A. Reynolds, personal communication, February 25, 2015).
- Children having one or two years of CPC preschool experience were less likely than those having no CPC preschool experience to have received special education throughout the elementary school years (Reynolds, 1995).

The expectation for the PFS-funded expansion of CPC to new sites in CPS and increasing the number of available CPC preschool slots at existing sites in CPS was based on previous research showing positive impacts on kindergarten readiness and school achievement, and reductions in special education placements over time.

Below we discuss the findings from the year 2 evaluation outcomes, including some of the limitations in interpreting the data.

Kindergarten Readiness Findings

To put the findings in context, we ask four questions of the kindergarten readiness data. First, is the assessment a reliable and valid measure of children's kindergarten measure and how much error might exist in measuring children's "true" kindergarten readiness? Second, to what extent are our findings similar to those of other CPC and CLS data? Third, to what extent are our findings similar to the ECLS-K data for the general population and for children from low-income families? Finally, what might explain the decrease in the percentage of children ready for kindergarten in the two cohorts of children from year 1 to year 2?

IS THE ASSESSMENT A RELIABLE AND VALID MEASURE OF CHILDREN'S KINDERGARTEN READINESS?

It is important to understand the current context of assessing young children's kindergarten readiness. Currently, there is no summative assessment tool that is considered a 'gold standard' for measuring kindergarten readiness. All assessments, whether they involve teacher ratings or direct assessments of young children's abilities and skills, are prone to some measurement error. Direct assessments are limited in that they only assess children at a single time point and often only assess a very small set of skills (e.g., counting, alphabet recognition) administered and scored based on performance on a highly defined set of items that do not reflect the full spectrum of foundational skills needed to be ready for kindergarten and succeed in school. Teacher ratings that are based on longer periods of observation across multiple settings can be a more accurate representation of children's skills and behaviors. The developers of the TS *GOLD*™ recommend certain kinds of trainings to help teachers use the tool as well as procedures to ensure the accuracy of ratings (e.g., checking inter-rater reliability).

Some aspects of the TS *GOLD*™ measure show strong psychometrics. For example, the developers have published analyses showing strong internal reliability (Cronbach's alpha reliability coefficients = .94 - .97) which suggests that the items are correlated (i.e., a child who scores high on one item has a high probability of scoring high on another item (Burts & Kim, 2014). There is limited evidence, however, for

test-retest or inter-rater reliability on the TS *GOLD*™. Understandably, high test-retest reliability may not be something you would expect from this measure as children develop and mature very rapidly in the first five years, especially children participating in a preschool program. However, we would want to have higher inter-rater reliability, meaning different raters (in this case, teachers) would tend to rate the children the same way given the same information and observation window for those two raters. There are limited data, however, on whether this measure has moderate to high interrater reliability. For example, correlations between the ratings of a master trainer and ratings of teachers who were current users of the system were high (between .80 and .90) suggesting that when teachers are adequately trained (in this case, received a full two days of training) they can demonstrate good reliability when making ratings of the same child (Burts & Kim, 2014).

There is some evidence that the measure has validity, meaning it measures what it is intended to measure. One way to assess the measure's validity is to examine whether children with a known delay or disability perform lower on the ratings when compared to children who do not have identified delays or disabilities. The developers have published data showing children with identified disabilities have lower TS *GOLD*™ scores when compared with their typically developing peers and they develop at a slower rate (Lambert, Kim, and Burts (2014b). A different way to examine validity is to test whether teacher ratings are correlated to independent direct assessments of children's development. The correlations between TS *GOLD*™ domain scores and independent direct assessment and other teacher rating measures are moderate (in the range of .3 to .5) suggesting they may be measuring different aspects of development (Decker, 2013; Soderberg et al., 2013; Teaching Strategies, 2013)(GOLD Concurrent Validity, 2013). In addition, to our knowledge, validation studies that examine predictive relationships have not been conducted (i.e., how well TS *GOLD*™ assessment ratings predict later academic achievement and behavior in kindergarten and beyond).

Given that our sample included many children who are ELLs, it is important to understand the validity of the use of the TS *GOLD*™ with that population. Kim, Lambert, and Burts (2013) published data that provide some empirical evidence supporting the validity for the TS *GOLD*™ domains and learning objectives for

typically developing children, as well as English-language learners and for those children identified with special needs or disabilities. In other words, this observation-based teacher rating assessment tool measures the construct domains in the same way across various subgroups of children 3 to 5 years old. These studies were conducted with samples of teachers who had been trained to reliability. For the measure to be used with children who have very limited English abilities, the teacher must be bilingual in both English and the child's native language and/or be able to document children's skills and abilities using information from a bilingual teacher assistant or family members. If the teacher is not bilingual and/or has not gathered additional documentation and observation data from other staff, it is possible that the teacher will either under- or over-estimate the child's abilities.

TO WHAT EXTENT ARE THE PFS-CPC KINDERGARTEN READINESS FINDINGS SIMILAR TO THOSE OF OTHER CPC AND CLS DATA?

More recently, Reynolds and colleagues (2014) published data in a peer-reviewed journal showing that 80.9% of children attending full-day CPC classrooms (n = 409) and 58.7% of children attending part-day CPC classrooms (n = 573) were considered kindergarten-ready when kindergarten-readiness was defined as meeting the national norm on four of the six TS GOLD™ domains. Additionally, full-day participants demonstrated higher average levels of skill mastery in the domains of language, mathematics, socio-emotional development, and physical health (but not for literacy and cognitive development). Reynolds and colleagues (2014) report a higher proportion of children who are kindergarten ready than that reported here, but use a less stringent standard for "readiness," i.e., a threshold of proficiency on four compared with five of the TS GOLD™ domains; five was the standard used for the current evaluation. If we had used that less stringent standard of 4 of 6 domains, an additional 9% would meet the kindergarten readiness criteria in Cohort 1, for a total of 70% and an additional 11% in Cohort 2 for a total of 53% (see Exhibit 6). Kindergarten readiness rates in cohorts 1 and 2 are similar or better than the previous CLS study findings for this outcome, but the percentage of children in Cohort 2 meeting kindergarten readiness benchmarks is lower than other contemporary studies of CPC and lower than the percentage found for Cohort 1. Potential explanations for these differences are discussed in more detail below.

TO WHAT EXTENT ARE PFS-CPC FINDINGS SIMILAR TO NATIONALLY REPRESENTATIVE STUDIES OF KINDERGARTEN READINESS AND FOR CHILDREN FROM LOW-INCOME FAMILIES?

Data from the contemporary, nationally representative sample of ECLS-K[23] children and using a single measure of kindergarten readiness which is similar to that of this report indicate rates of school or kindergarten readiness that are typically less than 50% for children from economically disadvantaged households (Isaacs, 2012). In comparison, the same report showed that 75% of children from more economically advantaged households (i.e., moderate to high income households) were considered ready for kindergarten. Other publicly available reports produced by states that are using a kindergarten readiness assessment show considerable variability given the different ways to measure kindergarten readiness. The state of Washington uses a modified version of the TS *GOLD*™ assessment and recently published data from their kindergarten entry assessment study showing that 31% of entering kindergarteners from low-income households met the benchmarks[24] on all six domains – what they refer to as "full" readiness (State of Washington, n.d.). About half of Cohort 1 (51%) and about one-fifth of Cohort 2 children (23%) demonstrated readiness in all domains (see Exhibit 6). About 72% of children (across income levels) met the benchmarks for at least 4 domains in Washington state (State of Washington, n.d.). In this evaluation, 70% of children in Cohort 1 and 53% of children in Cohort 2 (all low-income) met the benchmarks in at least 4 of the 6 domains. Thus, the rates of kindergarten readiness found for Cohort 1 and Cohort 2 are similar to what has been found in national and state studies.

WHAT EXPLAINS THE DECREASE IN KINDERGARTEN READINESS IN THE TWO COHORTS OF CHILDREN FROM YEAR 1 TO YEAR 2?

There is much larger year-to-year variation than expected in the percentages of children identified as meeting the kindergarten readiness criterion in Cohort 1 versus Cohort 2. That is, one would expect the rates to be similar in the two cohorts if three factors were similar from year to year: schools were serving the same or very similar populations of children; all teachers in both cohorts received the same type and quality of training and reliability checks to reliably complete the TS *GOLD*™ ratings

[23] ECLS-K is a contemporary longitudinal dataset that draws from a nationally representative sample; includes direct assessments of children's skills at kindergarten entry (cf. Hair, Halle, Terry-Humen, Lavelle, & Calkins, 2006; Lee, Zhai, Brooks-Gunn, Han, & Waldfogel, 2014).

[24] It is unclear what benchmarks on the TS *GOLD*™ assessment the Washington study uses.

(including new teachers who might need more training and practice in rating children reliably); and there were no major programmatic changes in how the CPC programs were implemented (e.g., instructional practices, curricula used, etc.) at each site from year 1 to year 2. If these were all similar from year 1 to year 2, then the only explanation for these differences would be that the assessment tool has a significant degree of unreliability. Finally, it is possible any combination of these *four* factors could explain the year-to-year variability, and different combinations of these factors might explain the year-to-year change at different sites.[25]

SRI conducted additional analyses to examine if and to what degree these factors might explain the decrease in percentage of children who are deemed kindergarten ready in Cohort 2 compared with Cohort 1.

Are the sites serving a different population?

There were some differences in the demographic characteristics served in Cohort 1 versus Cohort 2. Most significantly, the number and percentage of children who are designated as English Language Learners increased from Cohort 1 to Cohort 2 (Exhibit 5), from 44% to 57%. There were a larger number of ELLs served in year 2 and it could be harder to reliably assess these children with the TS GOLD™. Some children who are ELLs could be rated on their English language and literacy abilities while other children could not. The TS GOLD™ developers have provided guidance on whether or not children's language and literacy skills can be rated in English.

We also found that across Cohorts 1 and 2, the percentage of children meeting the Kindergarten readiness benchmark differs significantly by ELL status: 54% of all non-ELLs meet the 5 or 6 criterion compared with 42% of all ELLs ($p < .001$). Since Cohort 2 has more ELLs than Cohort 1, that difference might partially explain the lower rate of kindergarten readiness in Cohort 2 compared with Cohort 1 as children who are learning two languages may take longer to demonstrate proficiency in any of these domains.[26]

[25] CPS is exploring these factors related to teacher training and program implementation.
[26] We did not examine fall to spring gains for ELLs and non-ELLs, so we do not know whether ELLS are making good growth on the skills that are rated but just do not gain enough to meet the kindergarten readiness benchmark that we set.

Did all teachers receive the same training and reliability checks?

Although SRI does not collect information on teacher training and turnover, we did examine statistically the percentage of variance in children's scores that can be attributed to child, teacher, and school factors which helps us understand if the level of teacher effects or site effects is greater than what we would expect on a teacher rating. On the TS *GOLD*™ assessment, children receive a score on each of six domains. The assumption is that these scores represents the child's ability in the area measured and that a child's ability would be measured the same no matter which teacher completed the assessment. To test if this assumption is true, we can estimate the degree to which any child's score can be predicted by the teacher who completed the TS *GOLD*™. From other studies using TS *GOLD*™*,* we expect the rater to predict about 29 to 35% of the variance in a group of children's scores (Lambert, Kim & Burts, 2014). In the current study, the rater predicted 32 to 53% of the variance in children's scores depending on the domain. This amount of variation is more than the expected amount. This measurement error reduces the reliability of the scores. As scores become less reliable, it is more difficult to estimate the impact of an intervention on the ability being measured. There should be consistency in how different teachers rate the children. You would expect children with the same abilities to be rated similarly by different teachers. These analyses might reflect teacher training and turnover factors and these are described in greater detail by the CPS/city of Chicago program teams in an addendum.

Did CPC implementation vary from year to year?

There was a surprisingly large range in the percentage of children deemed ready for kindergarten across sites and across cohorts (24% to 71%). One factor that might influence these site differences could be site-level difference in implementation of the CPC model. For example, unobserved site factors may influence training in the CPC model, the ways that the CPC program is implemented, and other program components that may influence teacher ratings and/or children's kindergarten readiness (e.g., changes instructional practices). While the evaluation did not collect any data on site implementation, these factors are being explored by the implementation partners, CPS and the city of Chicago, who are conducting additional analyses about the contextual site factors that may have impacted implementation.

Special Education Placement Findings

Based on the extensive CLS analyses as well as reviewing existing data for CPS in the years leading up to the PFS-CPC project, the PFS project was built on the hypothesis that high-quality preschool through the CPC program would help prevent or reduce the need for special education in the intervention group. We found that the special education placement rate in kindergarten for the Cohort 1 intervention group was slightly lower than the rate for the comparison group (4.38% versus 4.94%). This is a difference of .56% in the rate of special education use between the two groups, and is an 11% decrease for the treatment group relative to the comparison group.

We want to acknowledge a few data limitations. As described earlier, SRI was unable to identify with certainty children from the comparison group who may have received pre-K programming in non-CPS funded settings in order to exclude them from the comparison group sample. We did however exclude from the comparison group all children who had a CPS identification number when they entered kindergarten, which serves as an indication that they attended a CPS preschool program and/or were receiving services from the school district. In addition, with the propensity score weighting procedure, we used a large comparison sample to reduce the impact of possible preschool attendance on our estimation of comparison group outcomes. That is, there is likely some proportion of children in the 9,445 comparison group sample who did attend some type of public preschool program (e.g., Head Start) or private child care program using child care subsidies and/or tuition. It is not clear how inclusion of some children with preschool experience in the comparison group is affecting the special education rate in that group. We also want to note that in the original design of this PFS project using the CPC model, the planning team decided that the targeted population was high risk children but not necessarily those with identified disabilities prior to participation in preschool. Therefore, children who were identified prior to PreK were excluded from the sample that is tracked over time.

Finally, the evaluation team used disability categories to define who had a mild versus a severe disability, based on the assumption that the intervention was only targeting prevention of the need for later special education services for children at-risk of a mild disability. Because the disability categories do not provide information on the severity of a child's disability or delay, the team used additional information about placement in special day classes and blended classrooms to identify children

as having more “severe” delays or disabilities.” While the information sources used are proxies for identifying those with “severe” delays or disabilitis who were excluded from the analyses, rather than direct measurements of the severity of a child’s disability, they were agreed upon with input from CPS as appropriate proxies to identify severity. We recognize that some children with developmental delay may eventually develop a more severe disorder and need services as could some of the children with speech and language delays and emotional disturbances.

To put the findings about rates of special education placement for Cohort 1 and comparison group samples into context, we searched for comparable data and organized our search around three questions. First, to what extent are the findings similar to those reported in the Chicago Longitudinal Study of the earlier CPC cohort? Second, how do kindergarten special education placement rates for the intervention and comparison groups compare with kindergarten special education placement rates in the Chicago Public Schools? Third, to what extent are the findings similar to the national data about special education placement rates for children who attend preschool versus those who do not?

TO WHAT EXTENT ARE THE FINDINGS SIMILAR TO THOSE REPORTED IN THE CHICAGO LONGITUDINAL STUDY OF THE EARLIER CPC COHORT?

Studies of the Chicago Longitudinal Study (CLS) cohort of children who attended CPC preschool in 1984 showed that special education placement rates were lower for CPC recipients compared with children who did not participate in CPC preschool. In one CLS analysis that looked at children 6 to 18 years old, CPC participants had an *average* special education placement rate of 14.4% compared with 24.6% for children in the comparison group (Reynolds et al., 2003). This rate was calculated by averaging rates across school years, and therefore is not directly comparable to the kindergarten special education rates described in this report. In another CLS analysis that compared average special education placement rates from school entry through eighth grade for children who had attended a CPC preschool with a comparison group, special education placement rates were 12.5% versus 18.4%, respectively (Conyers et al., 2003). Significant differences in special education placement rates between children in the CLS who did and did not attend CPC preschool programs emerged as early as first grade (0.5% versus 3.2%) (Conyers et al., 2003). These earlier CLS findings suggest that special education rates may rise over the early

elementary grades and it is still possible to detect a positive intervention effect for reduced special education placement from attendance in CPC preschool but it may take time to detect this effect.[27] Other changes in how children are identified and placed in special education have occurred since the 1990s when these latter analyses were completed making these comparisons less appropriate than more contemporary data. These changes include but are not limited to improvements in early screening and programming prior to entering kindergarten as well as during the early elementary school years). Such improvements may mean that contemporary children who would have needed special education in the past are being identified during preschool or the early elementary grades and given more intensive or focused supports that can prevent those early difficulties from becoming more severe delays and learning problems requiring special edcuaiton services.

The percentage of children needing special education for developmental delay or speech/language issues was approximately 4% in both groups. Interestingly, the majority of children in the intervention cohort were identified during the preschool year while enrolled in CPC preschool classrooms. Because of changes in how children are identified for services and how important assessment and early intervention are in early care and education settings, we believe as part of best practices in a high-quality preschool classroom, more children may be identified for services (and appropriately so) early in their school careers than may have occurred previously when the CLS was conducted. The hypothesized role of CPC in preventing or reducing the need for special education is still valid; however, in contemporary settings, there may be an initial increase in identification and placement before we see long-term reductions and/or average reductions in this outcome.

HOW DO KINDERGARTEN SPECIAL EDUCATION PLACEMENT RATES FOR THE INTERVENTION AND COMPARISON GROUPS COMPARE WITH KINDERGARTEN SPECIAL EDUCATION PLACEMENT RATES IN CHICAGO PUBLIC SCHOOLS?

In the 2015–16 school year, the kindergarten special education placement rate for CPS overall for what we are identifying as *mild delays and disorders* was 7.4%. Our findings show that both Cohort 1 and their comparison group have rates that are

[27] Support for this suggestion comes from earlier CPC studies that found that students with extended *CPC program participation (through second or third grade) had lower rates of special education* placements than CPC students having fewer years of intervention through middle (Reynolds & Temple, 1998) and high school (Reynolds, Temple, Robertson, & Mann, 2001).

lower than the rate in CPS overall during the same year. We do not have a good explanation for this difference. Historical data from CPS demonstrate that special education placement rates during preschool have been relatively high for children attending CPS preschool (10–14% for the last 5 school years)[28] and that the overall CPS rates decrease in kindergarten (7–8%) and steadily climb through elementary and middle school grades; by third grade, for example, overall special education rates in CPS have been 10–11% for the last 5 school years (Chicago Public Schools, 2016). These special education rates in CPS suggest that we will see increases in special education rates as children move through the early elementary years, probably for both groups. But we hypothesize that the rate for the intervention group will increase at a slower rate and be lower than the rate in the comparison group in the later elementary years if the preschool CPC experiences have helped intervention group children have better developmental trajectories. Continued follow-up of the cohorts will address this question.[29]

TO WHAT EXTENT ARE THE FINDINGS SIMILAR TO THE NATIONAL DATA ABOUT SPECIAL EDUCATION PLACEMENT RATES FOR CHILDREN WHO ATTEND PRESCHOOL VERSUS THOSE WHO DO NOT?

Data collected annually from every state as part of federal reporting required under the Individual with Disabilities Education Act by the Office of Special Education Programs give us another perspective on the special education rates. These national data indicate that 7% of children were receiving special education services at age 5, the kindergarten year for most students (National Center for Education Statistics, 2007). Other data from the nationally representative Head Start Family and Child Experiences Survey (FACES) for children participating in Head Start preschool programs (Barton, Spiker, & Williamson, 2012) and the nationally representative ECLS-B study (Parsons, 2016)[30] found rates higher than those for our comparison sample (8% and 7%, respectively). However, findings from the FACES study also suggest that more children in Head Start without an Individual Education Plan (IEP) meet other criteria for disability or delay than are served (about 33% meet criteria

[28] For 2015-16 school year, we found that for low-income children who had attended any CPS preschool, the overall IEP special education rate in kindergarten was 13.7% (10.7% for mild and 3% for not mild delays and disorders).
[29] The special education rates in CPS preschool may be higher than rates for kindergarten because children identified with a delay or disability prior to kindergarten would be likely to be referred to CPS and when determined to be eligible for an IEP, they would be served in a CPS preschool program.
[30] ECLS-B children entered kindergarten in 2006 or 2007.

indicative of delays such as very low assessment scores on standard measures, but only 8% had an IEP and were receiving special education services) (Barton et al., 2012). These national data and other research suggest that children from low-SES families are at greater risk of developmental delay and low levels of kindergarten readiness but may be less likely than higher SES children to be receiving special education services in early childhood (Morgan, Farkas, Hillemeier, & Maczuga, 2012). Taken together, these other data suggest that the rates we are seeing in both the intervention and comparison groups are lower that would be expected for our high risk samples.

Limitations

The evaluation is limited to data already collected in the district data and as such, must use for a kindergarten readiness measure, an assessment tool that was not developed for these purposes. The TS *GOLD*™ assessment is vulnerable to large (and maybe small) changes in training protocols and teacher turnover and both reliability and validity may suffer. In addition, the use of the TS *GOLD*™ measure with ELL populations is problematic because some ELL children can be assessed on all six domains in English while others cannot. Furthermore, there are not scale scores for children who have the TS *GOLD*™ language and literacy assessment domains in Spanish and therefore we did not have complete data needed for making the kindergarten readiness determination for those children.

Use of a teacher-report measure as the indicator of kindergarten readiness can be seen as a limitation of this study. As described above, teacher effects or bias is also a limitation in the evaluation. The accuracy of teacher ratings of young children's behavior has been questioned by some early childhood researchers (Baker et al, 2015; Mashburn & Henry, 2004). Additionally, there is a need in the field for more research about the amount and types of teacher training and knowledge needed to assess young children reliably and accurately, even while it is widely acknowledged that teacher training and knowledge are critical (Institute of Medicine & National Research Council, 2015; National Association Early Childhood Specialists in State Departments of Education & National Association for the Education of Young Children, 2003; National Association for the Education of Young Children, 2009).

For the special education outcome at kindergarten, to the best of our ability we created a comparison group that was weighted to match closely the characteristics of children in the intervention cohort. However, both propensity score methods–PSW and PSM–have some disadvantages. One disadvantage is that these methods only account for observed (and observable) covariates. They cannot balance intervention and comparison groups on unobservable characteristics (for example, parent education levels or parent involvement in the children's early learning). Second, we can only identify those children who attended a CPS-funded preschool program prior to kindergarten. Many entering kindergarten children who meet the income and age criteria may have attended other preschool programs either public (e.g., state-funded preschool outside of the city of Chicago) or private (e.g., using child care subsidies for example). The evaluation team had no way of identifying which children attended these other settings because this information is not routinely collected at kindergarten entry in CPS.

Conclusion

Together, these findings show that early childhood education in the form of CPC is associated with rates of kindergarten readiness of 61% in Cohort 1 and 42% in Cohort 2. The findings also show that the decrease in special education in kindergarten is small and that overall special education rates in both the intervention and comparison group children are very low at the end of kindergarten.

Next Steps

The year 3 report will include kindergarten readiness outcomes for a new group of children in Cohort 3. It will also present data on special education placement rates at the end of kindergarten for Cohort 2 children compared with rates of special education placement in a matched comparison sample of children who did not attend any preschool in CPS. In addition, we will present the continued follow-up data on special education placement rates at the end of first grade for Cohort 1 and its comparison group.

References

Baker, C. N., Tichovolsky, M. H., Kupersmidt, J. B., Voegler-Lee, M. E., & Arnold, D. H. (2015). Teacher (mis)perceptions of preschoolers' academic skills: Predictors and associations with longitudinal outcomes. *Journal of Educational Psychology, 107*(3), 805-820. doi:10.1037/edu0000008

Barton, L., Spiker, D., & Williamson, C. (2012). Characterizing disability in Head Start programs: Not so clearcut. *Early Childhood Research Quarterly, 27*(4), 569-612.

Becker, S. O., & Inchino, A. (2002). Estimation of average treatment effects based on propensity scores. *Stata Journal, 2*, 358-377.

Belfield, C., Nores, M., Barnett, W. S., & Schweinhart, L. J. (2006). The High/Scope Perry Preschool Program: Cost-benefit analysis using data from the age-40 followup. *Journal of Human Resources, 41*(1), 162-190.

Blackorby, J., Schiller, E., Mallik, S., Hebbeler, K., Huang, T., Javitz, H., Marder, C., Nagle, K., Shaver, D., Wagner, M., & Williamson, C. (2010). *Patterns in the identification of and outcomes for children and youth with disabilities*. Menlo Park, CA: SRI International.

Burger, K. (2010). How does early childhood care and education affect cognitive development? An international review of the effects of early interventions for children from different social backgrounds. *Early Childhood Research Quarterly, 25*(2), 140-165.

Burts, D. C., & Kim, D.-H. (2014). The Teaching Strategies GOLD Assessment System: Measurement Properties and Use. *NHSA Dialog, 17*(3).

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys, 22*(1), 31-72.

Campbell, F. A., Ramey, C. T., Pungello, E., Sparling, J., & Miller-Johnson, S. (2002). Early childhood education: Young adult outcomes from the Abecedarian Project. *Applied Developmental Science, 6*(1), 42-57.

Chicago Public Schools. (2016). School data: Chicago Public Schools.

Conyers, L. M., Reynolds, A. J., & Ou, S. (2003). The effect of early childhood intervention and subsequent special education services: Findings from the Chicago Child-Parent Centers. *Educational Evaluation and Policy Analysis, 25*, 75-95.

Curtis, L. H., Hammill, B. G., Eisenstein, E. L., Kramer, J. M., & Anstrom, K. J. (2007). Using inverse probability-weighted estimators in comparative effectiveness. Analyses with observational databases. *Medical Care, 45*(2), 158-165.

Decker, C. G. (2013). *Teaching Strategies GOLD: Testing reliability and validity using the Bracken School Readiness Assessment*. Unpublished report of CAP Tulsa.

Gaylor, E., Kutaka, T., Ferguson, K., Williamson, C., Wei, X., & Spiker, D. (2016). *Evaluation of kindergarten readiness in five Child-Parent Centers: Report for 2014-15*. Menlo Park, CA: SRI International.

Hair, E. C., Halle, T., Terry-Humen, E., Lavelle, B., & Calkins, J. (2006). Children's school readiness in the ECLS-K: Predictions to academic, health, and social outcomes in first grade. *Early Childhood Research Quarterly, 21*, 431-454.

Harder, V. S., Stuart, E. A., & Anthony, J. C. (2010). Propensity score techniques and the assessment of measured covariate balance to test causal associations in psychological research. *Psychological Methods, 15*(3), 234-249.

Harding, D. (2003). Counterfactual Models of Neighborhood Effects: The Effect of Neighborhood Poverty on Dropping Out and Teenage Pregnancy. *American Journal of Sociology, 109*(3), 676-719. doi:10.1086/379217

Heckman, J., Ichimura, H., Smith, J., & Todd, P. (1998). Characterizing selection bias using experimental data. *Econometrica, 66*(5), 1017-1098. doi:10.2307/2999630

Hirano, K., Imbens, G. W., & Ridder, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica, 71*(4), 1161-1189. doi:10.1023/a:1020371312283

Human Capital Research Collaborative. (2015). *2015-16 program guidelines and requirements. CPC-PK3 program*. Minneapolis, MN: University of Minnesota. Retrieved from https://humancapitalrc.org/midwest-cpc/cpc-resources

Institute for Research on Poverty. (1997). Investing in young children. *Focus, 19*(1), 1-4.

Institute of Medicine, & National Research Council. (2015). *Transforming the workforce for children birth through age 8: A unifying foundation*. Washington, DC: The National Academic Press.

Isaacs, J. B. (2012). *Starting school at a disadvantage: The school readiness of poor children*. Washington DC: Center on Children and Families at Brookings.

Karoly, L. A., Greenwood, P. W., Everingham, S. S., Hoube, J., Kilburn, M. R., Rydell, C. P., Sanders, M., & Chiesa, J. (1998). *Investing in our children: What we know and don't know about the costs and benefits of early childhood interventions* (MR-898-TCWF). Santa Monica, CA: RAND Corporation.

Kilday, C. R., Kinzie, M. B., Mashburn, A. J., & Whittaker, J. V. (2012). Accuracy of teacher judgments of preschoolers' math skills. *Journal of Psychoeducational Assessment, 30*, 148-159.

Kim, D. H., Lambert, R. G., & Burts, D. C. (2013). Evidence of the validity of Teaching Strategies GOLD® assessment tool for English language learners and children with disabilities. *Early Education & Development, 24*(4), 574-595.

Knudsen, E. I., Heckman, J. J., Cameron, J. L., & Shonkoff, J. P. (2006). Economic, neurobiological, and behavioral perspectives on building America's future workforce. *Proceedings of the National Academy of Sciences, 103*(27), 10155–10162.

Lambert, R. G., Kim, D. H., & Burts, D. C. (2014a). *Technical manual for the Teaching Strategies GOLD® assessment system (3rd edition)*. Center for Educational Measurement and Evaluation, UNCC, Charlotte.

Lambert, R. G., Kim, D. H., & Burts, D. C. (2014b). Using teacher ratings to track the growth and development of young children using the Teaching Strategies GOLD® assessment system. *Journal of Psychoeducational Assessment, 32*(1), 27-39.

Lechner, M. (2002). Some practical issues in the evaluation of heterogeneous labour market programmes by matching methods. *Journal of the Royal Statistical Society Series a-Statistics in Society, 165*, 59-82. Article. doi:10.1111/1467-985X.0asp2

Lee, R., Zhai, F., Brooks-Gunn, J., Han, W.-J., & Waldfogel, J. (2014). Head Start participation and school readiness: Evidence fro the Early Childhood Longitudinal Study-Birth cohort December 10, 2012. *Developmental Psychology, 50*(1), 202-215.

Lynch, R. G. (2007). *Enriching children, enriching the nation: Public investment in high-quality prekindergarten*. Washington, DC: Economic Policy Institute.

Mashburn, A. J., & Henry, G. T. (2004). Assessing school readiness: Validity and bias in preschool and kindergarten teachers' ratings. *Educational Measurement: Issues and Practice, 23*(4), 16-30.

Morgan, P. L., Farkas, G., Hillemeier, M. M., & Maczuga, S. (2012). Are minority children disproportionately represented in early intervention and early childhood special education? *Educational Researcher, 41*(9), 339-351.

Muschkin, C. G., Ladd, H. F., & Dodge, K. A. (2015). Impact of North Carolina's early childhood initiatives on special education placements in third grade. *Educational Evaluation and Policy Analysis*. Advance online publication. doi:10.3102/0162373714559096

National Association Early Childhood Specialists in State Departments of Education, & National Association for the Education of Young Children. (2003). *Early childhood curriculum, assessment, and program evaluation: Building an effective, accountable system in programs for children birth through age 8* (Position Statement with expanded resources). Washington, DC: National Association Early Childhood Specialists in State Departments of Education (NAECS/SDE), National Association for the Education of Young Children. Retrieved from www.naeyc.org/resources/position_statements/pscape.pdf

National Association for the Education of Young Children. (2009). *Developmentally approrpriate practice in early childhood programs serving children from birth through age 8* (Position Statement). Washington, DC: Author.

National Center for Education Statistics. (2007). *Timing and duration of student particpation in special education in the primary grades*. Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

National Research Council. (2008). *Early childhood assessment: Why, what, and how?* Committee on Developmental Outcomes and Assessments for Young Children, C. E. Snow & S. B. Van Hemel, (Eds.). Washington, DC: The National Academies Press.

Ravallion, M. (2001). The mystery of the vanishing benefits: An introduction to impact evaluation. *The World Bank Economic Review, 15*(1), 115-140. doi:10.1093/wber/15.1.115

Reynolds, A. J. (1994). Effects of a preschool plus follow-on intervention for children at risk. *Developmental Psychology, 30*, 787-804.

Reynolds, A. J. (1995). One year of preschool intervention or two: Does it matter? *Early Childhood Research Quarterly, 10*, 1–31.

Reynolds, A. J. (2000). *Success in early intervention: The Chicago Child-Parent Centers*. Lincoln, NE: University of Nebraska Press.

Reynolds, A. J. (2002). *Chicago's preschool programs: Do they promote children's school success? Executive Summary II*. Chicago, IL: Chicago Board of Education.

Reynolds, A. J., & Hayakawa, C. M. (2011). Why the Child-Parent Center education program promotes life-course development. In E. Zigler, W. S. Gilliam, & W. S. Barnett (Eds.), *The pre-k debates. Current controversies and issues*. New York, NY: Brookes.

Reynolds, A. J., Magnuson, K. A., & Ou, S. R. (2010). Preschool-to-third grade programs and practices: A review of research. *Children and Youth Services Review, 32*(8), 1121-1131.

Reynolds, A. J., Richardson, B. A., Hayakawa, M., Englund, M. M., & Ou, S.-R. (2016). Multi-site expansion of an early childhood intervention and school readiness. *Pediatrics, 138*(1), 1-11.

Reynolds, A. J., Rolnick, A. J., & Temple, J. A. (Eds.). (2014). *Health and education in early childhood: predictors, interventions, and policies*. Cambridge, United Kingdom: Cambridge University Press.

Reynolds, A. J., & Temple, J. A. (1998). Extended early childhood intervention and school achievement: Age thirteen findings from the Chicago longitudinal study. *Child Development, 69*, 231-246.

Reynolds, A. J., & Temple, J. A. (2008). Cost-effective early childhood development programs from preschool to third grade. *Annual Review of Clinical Psychology, 4*, 109-139.

Reynolds, A. J., Temple, J. A., & Ou, S.-R. (2003). School-based early intervention and child well-being in the Chicago Longitudinal Study. *Child Welfare, 82*(5), 633-656.

Reynolds, A. J., Temple, J. A., Robertson, D. L., & Mann, E. A. (2001). Long-term effects of an early childhood intervention on educational achievement and juvenile arrest: A 15-year follow-up of low-income children in public schools. *Journal of the American Medical Association, 285*(18), 2339-2346.

Reynolds, A. J., Temple, J. A., Robertson, D. L., & Mann, E. A. (2002). Age 21 cost-benefit analysis of the Title I Chicago Child-Parent Centers. *Educational Evaluation and Policy Analysis, 24*(4), 267-303.

Root, E. D., & Humphrey, J. L. (2014). Neighborhood racial composition and trajectories of child self-rated health: An application of longitudinal propensity scores. *Social Science & Medicine, 120*, 31-39. doi:http://dx.doi.org/10.1016/j.socscimed.2014.09.006

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41–55.

Rosenbaum, P. R., & Rubin, D. B. (1984). Reducing bias in observational studies using subclassification n the propensity score. *Journal of the American Statistical Association, 79*, 516-524.

Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *American Statistician, 39*, 33-38.

Sampson, R. J., Sharkey, P., & Raudenbush, S. W. (2008). Durable Effects of Concentrated Disadvantage on Verbal Ability Among African-American Children. *Proceedings of the National Academy of Sciences, 105*, 845-852.

Shonkoff, J., & Phillips, D. (Eds.). (2000). *From neurons to neighborhoods: The science of early childhood development*. Washington, DC: National Academy Press.

Soderberg, J. S., Stull, S., Cummings, K., Nolen, E., McCutchen, D., & Joseph, G. (2013). *Inter-rater reliability and concurrent validity study of the Washington Kindergarten Inventory of Developing Skills (WaKIDS)*. Unpublished report prepared for the State of Washington Office of Superintendent of Public Instruction.

State of Washington, Office of Superintendent of Public Instruction (n.d.). *Washington State report card. Washington kindergarten inventory of developing skills (WaKIDs)*. Olympia, WA: Author. Retrieved from http://reportcard.ospi.k12.wa.us/WaKidsDetailPage.aspx?domain=WaKIDS&year=2014-15&wakidsyr=201314&schoolId=1&waslCategory=1&numberOrChart=1&yrs=2014-15%29

Teaching Strategies. (2013). *Teaching Strategies GOLD assessment system. Concurrent validity*. Bethesda, MD: Author. Retrieved from https://teachingstrategies.com/wp-content/uploads/2017/03/GOLD-Concurrent-Validity-2013.pdf

ten Bensel, T., Gibbs, B., & Lytle, R. (2015). A Propensity Score Approach towards Assessing Neighborhood Risk of Parole Revocation. *American Journal of Criminal Justice, 40*(2), 377-398. doi:10.1007/s12103-014-9269-z

Appendices

Appendix A: Chicago Child-Parent Center Social Impact Bond Evaluation Plan

Appendix B: Timing of Cohorts

Appendix C: Propensity Score Matching/Weighting Process

Appendix A: Chicago Child-Parent Center Social Impact Bond Evaluation Plan

Chicago Child-Parent Center Social Impact Bond
Evaluation Plan

December 2, 2014

Table of Contents

INTRODUCTION AND STUDY OBJECTIVES

The purpose of this document is to describe the methodology to be used to evaluate the impact of the Child Parent Center (CPC) Social Impact Bond (SIB) expansion on three primary impact outcomes: Special Education Utilization, Kindergarten Readiness, and Third Grade Literacy. This document also describes additional research questions that the Evaluator will seek to explore in collaboration with CPS to help the CPCs improve their performance. This methodology will be developed in conjunction with CPS and other experts in the early education field.

Participants in the CPC program (the Treatment Group) will be compared to groups of matched comparison students who did not have a CPC experience through the use of a propensity score matching algorithm. One comparison group will consist of children who did not attend any form of CPS Pre-K (No Pre-K comparison group). Another group will consist of children who attended some other type of CPS pre-K program, such as Head Start or Pre-School for All (Other Pre-K comparison group).

Payments based on Special Education utilization for the SIB project will be calculated using the difference in outcomes between the Treatment group and the No Pre-K comparison group.

Payments based on Kindergarten Readiness and Third Grade literacy will be calculated using outcomes of the treatment group relative to national standards

The Other CPS Pre-K comparison group will be used for sensitivity analyses and for addressing other research questions not related to payment triggers.

For the purposes of calculating payments owed as part of the SIB transaction, impacts will estimated using the total population of eligible students at SIB CPC sites, and then scaled to reflect the actual number of seats funded by the Lenders. We will adjust the scaling factors annually to reflect observed mobility trends.

The primary impact outcome questions are as follows:

1. What is the impact of the CPC program on the rate at which students need an IEP?
2. What is the impact of the CPC program on Kindergarten Readiness as defined by performance on the TS Gold instrument (completed by teachers at the end of preschool)?
3. What is the impact of the CPC program on Third Grade literacy as defined by performance on the CPS 3rd grade assessment?

In addition to these impact outcome questions, this evaluation will also seek to answer qualitative research questions that will help improve the performance of the program going forward unrelated to the Pay for Success calculations. These research questions will be developed more fully in conjunction with CPS and other experts in the early education field, and will only be pursued subject to additional external funding. The questions may include:

1. How do the primary impact outcomes vary by key subgroups, including gender, race, prior pre-school attendance, English language learner status, and potentially other subgroups?
2. How is the CPC program impacting attendance in Pre-K? How does attendance vary by site? How does attendance vary compare to other CPS Pre-K programs? Are there policies in place at specific sites that could be driving improved attendance?
3. How does the CPC program support a transition to Kindergarten? What sites are better at retaining children from Pre-K to K, both within their host school and within the entire district? Where do children who transfer within CPS go and why? Are there different impact outcomes for students who have less mobility?
4. How successful is the CPC program at improving social-emotional learning outcomes (defined by the social-emotional components of the TS Gold instrument) compared to children enrolled in other CPS pre-K programs?
5. How successful is the CPC model at engaging parents? What strategies are the most effective at encouraging parental engagement? What strategies appear to have the greatest impact on children's outcomes?

This document will serve as a template for how the evaluation will be conducted. The Evaluator will draft a final Evaluation Plan to be approved by CPS, the City, the Project Coordinator with Approval of the Lender Committee (such term being defined herein as such term is defined in the Loan Documents of the Lenders) using this document as a framework. No changes to payment terms or payment terminology will be made.

STUDY POPULATION

Eligible Population – Treatment Group

The Treatment Group in this study will consist of four-year-olds[1] who are attending Pre-K at any of the CPC SIB sites, in full day or half day programs, who at any point during the school year are eligible for the National School Lunch Program (NSLP).

[1] The intention is to identify children in the "age cycle four" year – the year prior to when they are planning to attend Kindergarten. At the time of the drafting of this document, this was defined by CPS as attaining age four on or before September 1st. This age identification protocol may be adapted as necessary to capture these children.

In the first year of the program, the following sites will be considered CPC SIB sites:

- De Diego
- Melody
- Peck
- Thomas
- Fiske
- Hanson Park

In the second year of the program, two additional sites, identified by CPS and approved by the City, will be added to the list of CPC SIB sites in addition to the sites listed above. If SIB funding in future years is used to add classrooms at additional schools as part of this project, those schools can be considered CPC SIB sites as well. If SIB funding is removed from one of the above sites, that site will no longer be considered a CPC SIB site.

A child may enter the program based on CPS age elibility criteria. For the 2014/15 school year, this entailed being age 4 as of September 1st.

All four-year-olds at CPC SIB sites, including children attending full-day classes, will be included in the treatment group, subject to the exclusions listed below.

In the first year of the program, we anticipate that 374 new slots for four-year-olds will be created through the SIB program. In the second year of the program, we anticipate that we will create an additional 408 new slots for four-year-olds in addition to maintaining funding for the original 374. In the third year of the program, we anticipate that we will maintain the 782 new slots that were created in years one and two. In the fourth year of the project, we expect to provide funding for at least 680 slots. Overall at CPC SIB sites, we anticipate that approximately 840 four-year-olds will be served per year once the program is operating at scale, with 782 of those positions funded by the SIB. The new slot amounts will be finalized prior to the launch of each new cohort.

Year 1 contingency for CPC Treatment Group

Due to the timing of the contracting, some of the new classrooms to be added in the 2014/15 school year will not be ready to serve children until the school year has already begun. Five of the Year 1 CPC SIB Sites where we will be adding additional classrooms (De Diego, Melody, Peck, Thomas, and Fiske) have been operating as a CPC for a year or more. As a result, they have an established leadership team, trained and experienced teachers, and fully outfitted classrooms.

To ensure that the children being tracked are receiving a sufficient dosage of the CPC program, for Year 1 only we will restrict the Treatment group eligibility to children who are enrolled in one of these five established CPC SIB sites, in a classroom that was already established as of September 2nd 2014 (the start of the 2014/15 school year). CPS will proceed with opening the new classrooms once all contractual issues have been resolved, but the children who are enrolled in those classrooms (including children at Hanson Park, the new CPC for Year 1) will not be included in the outcome calculations for the purposes of determining payments. This will allow CPS leeway to identify and train high quality teachers, and mitigate the risk that the outcomes (or underlying characteristics) of children who enroll in a CPC Pre-K after the start of the year are different from those of their peers who enrolled at the start of the year. The outcomes of these late-enrollees can be used as a unique sub-group, but will not factor into any calculations that determine payment amounts.

It is anticipated that the sample size of eligible four-year-olds in existing classrooms at existing CPC SIB sites will be at least 300 students. As with future analyses, when calculating payments this number will be scaled to reflect the actual number of slots funded by the Lenders as part of this initiative.

Eligible Population – No Pre-K Comparison Group

The No Pre-K Comparison Group in this study will be identified via a propensity score matching algorithm that pulls from a pool of eligible No Pre-K children districtwide. The pool of eligible No Pre-K children will include all children who meet the following criteria:

- Are enrolled in a CPS Kindergarten program, excluding:
    - Charter schools
    - Schools currently operating a CPC, as part of the SIB program or otherwise
    - Magnet and Selective Enrollment Schools
    - Schools that serve exclusively a special education population
- Are five years of age as of September 1st
- Did not attend a CPS Pre-K program in the school year prior to beginning Kindergarten
- Did not attend a Head Start program funded through the City of Chicago
- Are eligible for NSLP at any point during the school year

A child will be considered to have attended a Pre-K program if that child attended 10 days or more of a city funded pre-school program, or any days at any CPC site over the course of the school year. Days need not have been attended consecutively.

The No Pre-K Comparison group will be identified the year that their matched Treatment cohort begins Kindergarten to ensure that children within both groups are on the same age cycle.

Eligible Population – Other CPS Pre-K Comparison Group

The Other CPS Pre-K Comparison Group in this study will be identified via a propensity score matching algorithm that pulls from a pool of eligible children who attended other forms of CPS pre-K within the district. The pool of eligible Other CPS Pre-K children will include children who meet the following criteria:

- Are enrolled in a CPS Pre-K program, excluding:
  - Charter schools
  - Schools currently operating a CPC, as part of the SIB program or otherwise
  - Magnet and Selective Enrollment Schools
  - Schools that serve exclusively a special education population
- Are four years of age as of September 1st.
- Are eligible for NSLP at any point during the school year

The Other CPS Pre-K Comparison group will be identified the same year that their matched Treatment cohort begins pre-school to ensure that children within both groups are on the same age cycle. This group will only be identified subject to available external funding

Exclusions for payment calculations

The hypothesis is that the CPC program will have the biggest impact on children who are deemed at risk for poor school performance and achievement, but who lack a severe or significant disability. Without additional support, many of these children may end up being diagnosed with a mild learning disability, emotional disturbance, or developmental delay (including speech/language impairment). For these children, additional support in the classroom and at home can help ensure that they stay on track developmentally with their peers, avoiding the need for years of special education services.

The same impact is not expected for children with severe disabilities (identified in preschool or at a later date), and it is also not expected that a preschool intervention would meet the needs of the child without the benefit special education services, nor would that be appropriate or within the parameters of a child's right to a free and appropriate education. To ensure that children have access to the supports they need based on a clinical evaluation, if a child at any point during the course of the study is diagnosed with a severe disability, he or she will be removed from the study group during the year that the disability is added to the child's IEP onward. The preliminary list of severe disabilities, with input from the Independent Evaluator, may be as follows:

- autism
- deaf-blindness
- deafness
- hearing impairment
- orthopedic impairment
- other health impairment

- traumatic brain injury
- visual impairment
- multiply disabled[2]
- intellectual disability
- students placed into self-contained classrooms for children with special needs

This list may be adapted at the discretion of the Evaluator with approval from CPS, the City, the Project Coordinator, and the Approval of the Lender Committee.

RECRUITMENT PROCEEDURES

Children are identified for enrollment under the *Chicago: Ready to Learn!* application process. A timeline of application, placement, registration, and enrollment of children for the 2014/15 school year is provided below; this will also serve as an illustrative plan for how the process will occur in future years:

	Action	Description
April and June 2014	*Chicago: Ready to Learn!* Application Rounds 1 & 2	Parents obtain information about potential programs through chicagoearlylearning.org, cps.edu/readytolearn and the Chicago: Ready to Learn! hotline. Parents apply at application centers across the city for preschool under two application rounds.[3] The first round is held during the month of March - April and the second round is held during the month of May-June. Parents can choose up to three schools.
May and July/August 2014	Placement	Parents are offered a placement in a school and/or are placed on a waiting list. Children placed in a preschool program or on a waiting list are put into schools' Program Management in

[2] Intended to represent students with multiple severe disabilities

[3] For a complete list of application centers, see http://cps.edu/Schools/EarlyChildhood/Documents/ApplicationSites_SY14_15.pdf or http://cps.edu/readytolearn.Every CPC also is capable of accepting applications directly.

		the CPS SIM IMPACT system.
June through September 2014	Registration	Parents accept or decline placement. Schools notify parents of registration dates and times. Schools indicate parents' acceptance or decline of placement in Program Management and move registered children into the classroom Homerooms for IMPACT. Teachers complete the registration packet with families for all new students. Clerks enter identifying additional information into the IMPACT system.
September 2014	Enrollment	Children are enrolled upon attendance on the first day of school.
September 2014 onward	Rolling enrollment	Schools continue to enroll students throughout the school year as slots open up due to attrition, new funding, etc. Staff conduct additional outreach in communities with lower than expected enrollment to help fill all the slots. This includes additional ad spots, flyers, and community events. These children will only be included for evaluation purposes if they meet the dosage and eligibility requirements outlined in this document.

INTERVENTION AND OUTCOMES

Defining the Intervention

The CPC SIB intervention will provide one year of half-day CPC Pre-K to four-year-olds at CPC SIB sites. The key components of the CPC model are as follows:

Effective Learning Experiences

- Offer Pre-K classes that are limited to 34 children for half-day classrooms (two sessions of 17 children each) and have a minimum of 2 teaching staff. Full day classrooms, if available, will be limited to 20 children per session.
- Provide highly qualified educational staff that will provide the classroom instruction and parent engagement activities. For example, classroom teachers are certified with a bachelor's degree (or higher). Overall, program staff must adhere to the requirements set forth by the CPS Talent office, in accordance with collective bargaining unit agreements, and state regulations. Any changes in CPS education and certification requirements will be complied with.
- Use data to drive instruction by effectively documenting the organization and implementation of instructional practices to monitor quality and adherence to the Program, which is completed by all Program staff where appropriate.
- Program staff meet with parents over the course of each school year to review their child's progress and discuss parent program opportunities with the Parent Resource Teacher (PRT).

Aligned Curriculum

- Implement a CPS District curriculum and formative assessment that is aligned to standards, domains of learning, assessments, and learning activities.
- Collaborate with the PRT and classroom teachers to ensure that opportunities to engage families in student learning are available, appropriate and aligned to the program and parents' needs.
- CPS and, most specifically, the Office of Early Childhood Education provides meaningful professional development and ongoing coaching and feedback for teachers, aides, and other staff members that facilitates high-quality instructional practices.

Parent Involvement and Engagement

- Engage a PRT and School-Community Representative (SCR) to work closely with the Head Teacher and Liaisons to maintain a consistently supportive parent program.
- Encourage parents to sign a CPC school-home agreement at the start of the school year outlining a plan for fostering learning at home and participating in CPC activities.
- Offer and engage families in monthly activities. PRTs create and distribute a monthly parent involvement calendar, and conduct parent/teacher conferences over the year to review progress in the parent program.
- Provide a resource room dedicated to parent and family activities through Kindergarten when possible.
- Provide culturally responsive learning opportunities for families that provide flexibility for families' needs and schedules.

Collaborative Leadership Team

- Engage a Program leadership team that includes the Head Teacher, Parent Resource Teacher, and School-Community Representative.
- Meet regularly, under the direction of the Principal to discuss operations and best practices within the CPC.
- Meet regularly, under the direction of the OECE Management Team, with staff from across sites to share challenges, experiences, and best practices and makes frequent on-site visits to monitor quality and effectiveness to the Program.
- Establish meaningful partnerships with community providers to strengthen service delivery and enlist local universities in training opportunities.

Continuity and Stability

- CPC Pre-K classrooms are co-located in the same building as Kindergarten classrooms, when possible, to promote familiarity and integration for students as they transition to Kindergarten.
- Provide a structure of communication, planning, and joint activities, under the direction of the principal, Leadership team and OECE Management Team, from Pre-K through the primary grades.
  Provide a part-time Kindergarten aide when funding is available to support the transition into Kindergarten.

Professional Development System

- Offer ongoing professional development opportunities on current trends and needs in early childhood education classrooms, through the Office of Early Childhood Education and the CPC leadership teams, including topics such as quality curriculum and instruction, data driven instruction, learning environment, social and emotional needs, and parent engagement.
- Meet regularly and create professional learning communities to review ways to support their instruction in the classroom and with other teachers.

Defining Sufficient Dosage

Enrollment and attendance fluctuate throughout the year, with substantial changes during the early weeks of the school year. As a result, some of the children who start the year in a given classroom may not be the same children who end the year in that classroom. This may be due to for a variety of reasons such as mobility, a change in parents' schedules/ability to bring their children to school, or admission to a closer/more desirable program off of a waitlist later in the school year.

To ensure that CPC SIB children and families are receiving a minimum sufficient dosage of the CPC program, we will restrict analyses to children who attend a certain minimum cutoff of days. The Evaluator will examine historical data from CPS and other districts to determine trends in

attendance and identify a cutoff that sufficiently indicates that a child has received enough of the program for us to expect to see an impact. We are temporarily placing this cutoff at 66% of school days in a given school year; children who attended fewer than 66% of days during their Pre-K year will be omitted from the primary analyses.

The Evaluator may add additional criteria based on an analysis of enrollment and attendance data with the approval of CPS, the City, and the Project Coordinator and Approval of the Lender Committee.

Similarly, for the No Pre-K Comparison group, we will limit the primary analysis sample to eligible No Pre-K children who attend at least 66% of school days in a given school year. If a child at any point during the Kindergarten year attends a school operating a CPC program, that child will be omitted from primary analyses.

Defining Primary Impact Outcomes

Special Education Utilization
The primary Special Education utilization outcome will be defined as a binary indicator of whether or not a student has a CPS-issued Individualized Education Plan (IEP) in a given year. This will be a data point provided as part of the regular data collection points by CPS. As described above, if a student has a diagnosis on his or her IEP of a severe disability, that student will be removed from the study pool for the primary analyses. This indicator will be collected annually ever year Kindergarten through 6th grade.

Kindergarten Readiness
CPS uses the Teaching Strategies Gold (TS Gold) instrument in all their Pre-K classrooms to track the development of children. Based on teacher observations, TS Gold measures the progress of children in domains such as socio-emotional, physical, language, literacy, and cognitive development.

The TS Gold instrument is utilized nationally in Head Start programs and some publicly-funded preschool programs. The primary outcome metric for Kindergarten Readiness will be the share of children which are performing at or above the national trends across at least five out of the following six domains: Literacy, Language, Math, Cognitive Development, Socio-Emotional, Physical health.

Third Grade Literacy
Currently, CPS is planning to adopt the PARCC standardized exam. Treatment group children will be measured relative to national percentile rankings on this test or the accepted District assessment administered for 3rd grade. In following with Lesnick et al (2010)[4], every child

[4] See http://www.chapinhall.org/sites/default/files/Reading_on_Grade_Level_111710.pdf

reading at or above the 25th percentile on the English Language Arts/Literacy portion of the spring sitting of the PARCC test will be deemed to be reading at grade level. Any child reading at or above the 75th percentile nationally will be deemed to be reading above grade level. Any child reading below the 25th percentile will be deemed to be reading below grade level.

At the time of drafting this analysis, the PARCC test has yet to be officially implemented in CPS schools. Given the uncertainty of performance on this test and how its outcomes will compare to past tests taken by CPS students, the evaluator may suggest amendments to the definition of reading "on grade level" that could include utilizing a different test or metric. Any modifications must be made prior to the first cohort starting Third Grade, and must be approved by CPS, the City, the Project Coordinator, and Approved by the Lender Committee.

Defining Performance Improvement Questions

The details of these questions will be developed in conjunction with CPS and other partners over the 2014/15 school year. These analyses will be specified in full prior to the start of any data collection or analyses. These analyses will not affect the methodology or results of the primary impact outcomes, and will only be pursued subject to additional philanthropic or other funding.

DATA COLLECTION

Student data

Student data will be provided to the Evaluator by CPS. Pursuant to the data sharing agreement[5], CPS will strip sensitive individual identifiers and replace them with an anonymous student ID. The key variables CPS will provide are:

- Student ID
- CPS School ID of school currently enrolled in
- Date of Birth (or birth month & year)
- Days attended to date
- IEP status
- IEP diagnoses
- Reported race
- Reported ethnicity
- Free/reduced price lunch eligibility
- ZIP code of residence
- Fall and Spring TS Gold scores (if applicable)
- Any available variables on parental education
- Other variables deemed appropriate by the Evaluator and CPS for the purposes of creating a better propensity score match

[5] This data sharing agreement will be included as an appendix to this plan pending negotiation and drafting between CPS and the Evaluator.

Data will be collected on an annual basis on the based on the last school day in June which is reported for accuracy in the beginning of July. This may be adjusted based on discussions between the Evaluator and CPS to reflect the earliest date that all the necessary data would be available.

Neighborhood data

The Evaluator will pull neighborhood data from publicly available census data, such as the American Community Survey 5-year averages, which break out characteristics by zip code. Neighborhood data include:

- Neighborhood % of population in poverty
- Neighborhood % of population that are single mothers
- Neighborhood % of population that is Black
- Neighborhood % of population that is Hispanic
- Neighborhood % of population employed
- Neighborhood crime statistics
- Neighborhood health indicators[6]

The Evaluator will update the neighborhood data file when creating a new cohort of matched groups.

School data

Data on school level characteristics will be provided by CPS, including:

- CPS School ID
- Total student body population
- % Free/RP lunch
- % Black
- % Hispanic
- School-wide attendance rate from the 2013/14 school year
- School Rating (Levels 1, 2, or 3) from the 2013/14 school year[7]

These data, except for attendance and the school rating, will be updated annually. Attendance and rating data from SY2013/14 (or the closest assessment prior to SY2013/14) will remain fixed to reflect the fact that the presence of a CPC may improve attendance and the school rating over time, which could affect the matching algorithm for later cohorts. The Evaluator may adjust this protocol if extraneous events such as school closures, new leadership, or expansive new programs are added at individual schools or system wide that could contribute to imbalanced matches.

Data Security

---

[6] Crime stats and health indicators subject to availability of data. It may be possible to pull data from a Chapin Hall neighborhood analysis. These covariates may be omitted if it proves too difficult or costly to obtain them.

[7] All these data are publicly available online at http://www.cps.edu/schools/find_a_school/pages/findaschool.aspx. School rating is based on the CPS Performance Policy which is used to rate CPS schools. A Level 1 rating is "excellent", a Level 2 rating is "good" and a Level 3 rating is "low".

A data sharing agreement between CPS and the Independent Evaluator will define the parameters for sharing data required under this agreement.

STUDY DESIGN & OVERVIEW OF ANALYSES

Propensity score Matching Protocol

Comparison group students will be selected using a propensity score matching technique. Individuals from the treatment group will be matched to up to two individuals from the No Pre-K Comparison group and up to two individuals from the Other CPS Pre-K Comparison group. Matching will be conducted with replacement to allow comparison individuals to be matched more than once.

To create the Treatment Group in school year *t*, the Evaluator will receive the data collected on the last day of June of school year *t* from CPS of all four-year-olds who attended a SIB CPC in school year *t* up to the date of the data collection. The data collected and shared will contain all the student data elements listed above. After screening for eligibility as described above and removing ineligible students from the sample, the Evaluator will use students' ZIP codes to merge on neighborhood data, and students' school IDs to merge on school characteristics. Neighborhood data will be collected from a reliable source such as Chapin Hall. This will create a de-identified student-level file that contains student-level characteristics, characteristics of that student's neighborhood of residence, and characteristics of that student's school.

To create the No CPS Pre-K pool to be used for matching to the Treatment cohort in school year *t*, the Evaluator will receive a data dump on the last day of June of school year $t+1$ from CPS of all five or six-year-olds who attended a CPS Kindergarten in school year $t+1$ up to the date of the data dump. The data dump will contain all the student data elements listed above. After screening for eligibility as described above and removing ineligible students from the sample, the Evaluator will use ZIP code data to merge on neighborhood data, and school ID data to merge on school characteristics.

To create the Other CPS Pre-K pool to be used for matching to the Treatment cohort in school year *t*, the Evaluator will receive a data dump on the last day of June of school year *t* from CPS of all four-year olds who attended a CPS Pre-K program other than CPC in school year *t* up to the date of the data dump. The data dump will contain all the student data elements listed above. After screening for eligibility as described above and removing ineligible students from the sample, the Evaluator will use ZIP code data to merge on neighborhood data, and school ID data to merge on school characteristics.

To create the matched No Pre-K Comparison group, the Evaluator will append the Treatment Group dataset and the No Pre-K Comparison pool dataset, creating an indicator to identify which children are members of the Treatment group. The Evaluator will then run a probit model using the treatment indicator as the dependent variable and the following variables as independent variables:

- Race binary indicators
- Ethnicity binary indicators

- Gender ("Male" binary indicator)
- Parental education (subject to availability)
- Language spoken at home binaries
- Neighborhood % poverty
- Neighborhood % single mothers
- Neighborhood % by race
- Neighborhood % by ethnicity
- Neighborhood % employed
- Neighborhood crime rates (subject to availability)
- Neighborhood health indicators (subject to availability)
- Total student population of school currently attending
- % Free/RP lunch at school currently attending
- Racial composition of school currently attending
- Ethnicity composition of school currently attending
- School-wide attendance rate from the 2013/14 school year
- School Rating binaries from the 2013/14 school year

Using the results of this model, the Evaluator will predict a propensity score based on a student's observed characteristics. This score effectively represents the likelihood that a child, given his individual, neighborhood, and school level characteristics, would be in the Treatment group.

The Evaluator will use a nearest-neighbor matching algorithm[8] to identify the two closest matches based on propensity score for each Treatment group observation, with replacement.

Individuals from either the Treatment group or Comparison pool who are not matched will be dropped.

The remaining students from the Comparison pool who were matched will become the No Pre-K Comparison group for the remainder of the study. Comparison group students will receive a frequency weight equal to the number of times they were matched. Note that as a result, the Comparison group should contain approximately two times as many unique individuals as the Treatment group.

The same protocol will be used to identify the Other CPS Pre-K Comparison group, replacing the No CPS Pre-K Comparison pool with the Other CPS Pre-K Comparison pool.

A unique set of comparison groups will be created for each Treatment cohort (see Appendix for a cohort timing chart).

Checking for covariate balance between groups

Once the comparison groups have been identified, the Evaluator will check for balance between the groups across matching demographics. The Evaluator will choose appropriate methods to check for balance, including but not limited to normalized differences and t-tests of mean values of covariates between groups. If the Evaluator determines that there is imbalance in covariates

[8] By way of example, see "nnmatch" stata command

between groups, the Evaluator may choose to pursue a Matching Methodology Remedy as described below. The decision to pursue a remedy will be at the discretion of the Evaluator, taking into account the fact that with many matching variables and a p-value cutoff of .05, approximately 1 in 20 variables could have a statistically significant difference by random chance alone. The evaluator will consider the magnitude of the difference and the relative importance of the unbalanced variable(s) in question, placing particular attention to the individual-level race and gender indicators, the home language indicators, the neighborhood poverty indicators, and the school rating indicators.

Matching Methodology Remedies

In the event that the Evaluator deems that the propensity score matching algorithm has produced an inadequate match, the Evaluator may make modifications to the matching methodology. This could include introducing a caliper to ensure that certain variables are matched to within a narrow range (or matched exactly), adding or subtracting additional covariates, increasing or decreasing the number of matches, or other techniques deemed rigorous and appropriate by the Evaluator.

The Evaluator may also explore utilizing a set of comparison schools to limit the comparison pool. In this methodology, the Evaluator would identify a set of comparison schools that match the SIB CPC sites, identifying one to three schools for each site. The Evaluator would use a similar propensity score matching protocol, using school level characteristics, to identify these schools. From those comparison schools, the Evaluator would then perform a student-level propensity score match using a comparable methodology to the one described above. The Evaluator will then check for covariate balance to see if this produces better match results.

Once the Evaluator identifies a suitable comparison group that they deem to be well-matched on covariates, the Evaluator will present the match results, describing any changes that were made to the matching algorithm, which must be approved by CPS, the City, the Project Coordinator and Approved by the Lender Committee. The Evaluator should endeavor to use a similar matching protocol from year to year.

Calculating mobility factor

The theory behind the financing component of the SIB project is that providing the upfront intervention of high quality Pre-K can produce savings to CPS downstream through reduced Special Education utilization among the students served. For CPS to realize these savings, however, those students must remain in the CPS school district. If a student leaves the district, CPS would realize no savings from the fact that the intervention may have helped that that student catch up to his peers and prevented him from acquiring an IEP.

As a result, the Evaluator will calculate a Mobility Factor for each cohort that will represent the share of the original cohort that is still enrolled in a CPS school in a given year. This will be used to adjust the payment amounts to better reflect savings realized by CPS.

To calculate mobility, every year Kindergarten through 6th grade the Evaluator will determine what share of the original children in a given group from the first year of observation are still

enrolled in any CPS school. To do this, every year the Evaluator will send CPS a list of all the student IDs of the original group. CPS will match these IDs to their current enrollment database to determine which students were enrolled in a CPS school at any point in that school year. CPS will then return a dataset to the Evaluator indicating which student IDs are enrolled in a CPS school that year. The Mobility Factor will be defined as:

$$1 - \frac{\#\ of\ original\ students\ currently\ enrolled\ in\ any\ CPS\ school}{\#\ of\ students\ originally\ enrolled\ in\ the\ group}$$

By way of example, assume 500 Treatment group students were identified for the 2014/15 cohort. In SY2015/16, the Evaluator sends a list of these student IDs to CPS, who informs the evaluator that 460 of them are still enrolled at a CPS school. The cumulative mobility for that year would be 1 - 460/500 = .08. In SY2016/17, the Evaluator sends the original list of student IDs to CPS again, who informs the evaluator that 440 of them are still enrolled at a CPS school. The cumulative mobility for SY2016/17 would be 1 – 440/500 = .12.

For grades 7th through 12th, the Evaluator will impute a marginal mobility rate by averaging the incremental annual increase in the Mobility Factor over the last three years.[9] Every year, the Evaluator will impute a new Mobility Factor based on the average imputed marginal mobility rate. See Appendix B for a full example using hypothetical data.

Calculating effect size for Special Education utilization

To calculate the impact on Special Education utilization, the Evaluator will calculate the Average Effect Size per Person, which will then be scaled to reflect the number of seats funded by the Lenders for the purposes of calculating payments. This will allow the Evaluator to utilize all the data available, increasing sample sizes and precision of estimates.

To calculate this, the Evaluator will use the following equation:

$$AESP_{i,t} = SPEDC_{,i,t} - SPEDT_{,i,t}$$

where $AESP_{i,t}$ is the Average Effect Size per Person for cohort i in year t, $SPED_{C,i,t}$ is equal to the average of a binary indicator of Special Education utilization among the No CPS Pre-K Comparison group for cohort i in year t and $SPED_{T,i,t}$ is the average of a binary indicator of Special Education utilization among the Treatment group for cohort i in year t. At the discretion of the Evaluator and with approval from CPS, the City, the Project Coordinator, and the Approval of the Lender Committee, the Evaluator may regression-adjust this estimate to help account for any differences in covariates between the Treatment group and the Comparison group.

[9] The Evaluator may revise the methodology for averaging the mobility rate if they determine that the current methodology includes a grade breakpoint year that could result in abnormally high mobility out of the district. This methodology must be finalized before the first cohort reaches 6th grade.

Special Education outcomes will be calculated annually every year Kindergarten through 6th grade. Outcomes will be calculated separately for each cohort. Based on conversations with special education experts and reviewing existing CPS data, we believe that the vast majority of children who have a disability will be identified by the end of 6th grade. As a result, after the 6th grade effect size has been calculated, we will average the effect size over the last three years (4th, 5th and 6th grades) and lock in that average rate for the purposes of calculating payments in grades 7th through 12th. This lock-in rate will be calculated separately for each Treatment cohort. The Evaluator may propose changes to this lock-in methodology in the event that the Evaluator determines that this methodology produces skewed results. Any modifications must be approved by CPS, the City, the Project Coordinator, and Approved by the Lender Committee.

Calculating payments for Special Education utilization

To determine the size of Special Education payments owed in a given year for a given treatment group cohort, the Evaluator will multiply the Special Education Average Effect Size per Person for such cohort by the base cohort size multiplied by the 1 minus the cumulative mobility rate for that year. This will determine the Total Number of Special Education Slots Avoided for a given cohort in a given year:

$$Total\ Number\ of\ Special\ Education\ Slots\ Avoided = AESPi,t * BCSi * (1 - MFi,t)$$

where $AESP_{i,t}$ is the Average Effect Size per Person for cohort i in year t, BCS_i is the base cohort size for cohort i, and $MF_{i,t}$ is the cumulative mobility rate for cohort i in year t.

The base cohort sizes are based on the number of seats actually funded by investors. It is anticipated that the base cohort sizes will be as follows[10]:

Cohort Year	Base Cohort Size
2014/15	374
2015/16	782
2016/17	782
2017/18	680

The Total Number of Special Education Slots Avoided will then be multiplied by the Annual Savings Rate to determine the Special Education Payments owed for a given cohort in a given year. Negative payments will be rounded to zero. The Annual Savings Rate starts at a base of $9,100 in 2015 and grows 1% annually. The table below provides the rates through 2030:

Year	Savings Rate

[10] Note that actual sample sizes used for calculating effect sizes may be larger or smaller than the number of seats funded.

2015	$9,100
2016	$9,191
2017	$9,283
2018	$9,376
2019	$9,469
2020	$9,564
2021	$9,660
2022	$9,756
2023	$9,854
2024	$9,953
2025	$10,052
2026	$10,153
2027	$10,254
2028	$10,357
2029	$10,460
2030	$10,565

If applicable, the Special Education Payments from each cohort will be summed to produce the Total Special Education Payment owed by CPS for that year. These calculations will be reported to the Project Coordinator for the purposes of triggering payments to the Project Coordinator to be used to repay the lenders.

Payments for Special Education will be made every year K – 12th for each Treatment cohort.

Calculating effect size for Kindergarten Readiness

As part of the annual data pull, the Evaluator will receive spring TS Gold scores for Treatment group students. TS Gold regularly publishes a set of averages that reflect how children have scored nationally on TS Gold assessment sub-categories, broken out by the time of the test and the age in months of the child. Students will be classified as "meeting the national norms" for a sub-category if they score at or above the national mean spring score for that category for children in their age band.[11] The Evaluator will use the most up to date tables available.

Every child who scores at or above the national norm on at least five of the six subcategories in spring of their four-year-old pre-school year will be deemed "Kindergarten Ready." To calculate the Kindergarten Readiness payment, the Evaluator will calculate the share of the Treatment group students deemed Kindergarten Ready. The Evaluator will then multiply this number by the base cohort size, multiplied by cumulative mobility from the Kindergarten year of a given cohort. This will determine the Total Number of Kindergarten Ready Children for a given cohort. The Evaluator will then multiply this number by the payment rate of $2,900 to determine the total Kindergarten Readiness payments owed by the City for that cohort.

[11] See tables 5-14 of https://www.k12.wa.us/assessment/pubdocs/GOLDTechnicalManual2ndEditionLambert2.pdf for a list of the score thresholds.

Calculating effect size for Third Grade Literacy

CPS is currently transitioning to the PARCC exam. As a result, the exact methodology for calculating Third Grade Literacy may have to be adapted pending observation of how the test is being administered, scored, etc. In particular, in the event that data suggests that fewer than 50% of students are scoring above the 25th percentile, the Evaluator will propose a new protocol or test for determining Third Grade Literacy that better captures the performance of students. The Evaluator will propose a final protocol for approval by CPS, the City, and the Project Coordinator with Approval of the Lender Committee prior to the start of the 2018/19 school year – the year the first cohort begins 3rd grade. A draft protocol is below:

As part of the annual data pull, the Evaluator will receive 3rd grade spring PARCC scores for Treatment group students. The PARCC test is administered nationally, and as a result the outcomes of Treatment students can be compared to national averages. Students will be classified as "reading at or above grade level" if they score at or above the 25th percentile on the English Language Arts/Literacy portions of the PARCC exam.

To calculate the Third Grade Literacy payment, the Evaluator will calculate the share of the Treatment group students deemed to be reading "at or above grade level". The Evaluator will then multiply this number by the base cohort size, multiplied by cumulative mobility from the Third Grade year of a given cohort. This will determine the Total Number of Third Grade Children Reading at Grade Level for a given cohort. The Evaluator will then multiply this number by the payment rate of $750 to determine the total Third Grade Literacy payments owed by the City for that cohort.

Investigating Highly Unexpected Outcomes

The results of this evaluation will govern the flow of millions of dollars of payments. While it is the full intention of all parties to accept the results of the evaluation, in the event that a highly irregular outcome is achieved, a mechanism must be in place to validate the findings and confirm that they are due to the impact of the program, and not a flaw in the analysis or evaluation design. The Evaluator will have complete discretion to decide if and when a validation of the findings may be necessary, but the following events will serve as guiding principles that could suggest that a validation may be warranted:

- The difference in Special Education Utilization rates between the Treatment group and No Pre-K comparison group is negative or not statistically different from zero (p-value <.05) for any cohort in any year after Kindergarten
- The No Pre-K comparison group Special Education Utilization rate is more than 2.5 times the Treatment group Special Education Utilization rate for any cohort in any year after Kindergarten
- An irregular pattern from one year to the next in Special Education utilization for a given group, defined as utilization shrinking by more than two percentage points for a given group, or increasing by more than seven percentage points
- A larger impact observed when comparing a Treatment group cohort to its corresponding Other CPS Pre-K Comparison group any year after 1st grade.

The Evaluator will determine the appropriate techniques and mechanisms to employ to confirm the cause of the irregularity, which could include handchecking code, checking for continued balance in the treatment and comparison groups, and looking for policy changes within specific schools or system-wide that could have affected outcomes.

If the Evaluator finds a mechanical error, the results will be recalculated using the correction. If the Evaluator finds a methodological flaw, the Evaluator may propose a remedy to the evaluation plan to mitigate the inconsistency in future years. However, the results will not be recalculated for that year or any other past year. Changes to the plan must be approved by CPS, the City, and the Project Coordinator, and Approved by the Lender Committee.

APPENDIX A: TIMING OF COHORTS

	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024
Cohort 1											
Treatment	Identified and enroll in CPC PK	K	1st	2nd	3rd	4th	5th	4th-6th avg. SPED & Mobility rates locked			
Other CPS PK Comparison	Identified & enroll in other CPS PK	K	1st	2nd	3rd	4th	5th	4th-6th avg. SPED & Mobility rates locked			
No CPS PK Comparison		Identified and enroll in CPS K	1st	2nd	3rd	4th	5th	4th-6th avg. SPED & Mobility rates locked			
Cohort 2											
Treatment		Identified and enroll in CPC PK	K	1st	2nd	3rd	4th	5th	4th-6th avg. SPED & Mobility rates locked		
Other CPS PK Comparison		Identified & enroll in other CPS PK	K	1st	2nd	3rd	4th	5th	4th-6th avg. SPED & Mobility rates locked		
No CPS PK Comparison			Identified and enroll in CPS K	1st	2nd	3rd	4th	5th	4th-6th avg. SPED & Mobility rates locked		
Cohort 3											
Treatment			Identified and enroll in CPC PK	K	1st	2nd	3rd	4th	5th	4th-6th avg. SPED & Mobility rates locked	
Other CPS PK Comparison			Identified & enroll in other CPS PK	K	1st	2nd	3rd	4th	5th	4th-6th avg. SPED & Mobility rates locked	
No CPS PK Comparison				Identified and enroll in CPS K	1st	2nd	3rd	4th	5th	4th-6th avg. SPED & Mobility rates locked	
Cohort 4											
Treatment				Identified and enroll in CPC PK	K	1st	2nd	3rd	4th	5th	4th-6th avg. SPED & Mobility rates locked
Other CPS PK Comparison				Identified & enroll in other CPS PK	K	1st	2nd	3rd	4th	5th	4th-6th avg. SPED & Mobility rates locked
No CPS PK Comparison					Identified and enroll in CPS K	1st	2nd	3rd	4th	5th	4th-6th avg. SPED & Mobility rates locked

APPENDIX B: SAMPLE MOBILITY CALCULATIONS USING SIMULATED DATA

Sample Mobility Calculations Using Hypothetical Data					
School Year	Grade	Original Enrollment	Students still enrolled at a CPS school	Cumulative Mobility	Marginal Mobility
2014	PK	500	460	.08	.08
2015	K	500	440	.12	.04
2016	1st	500	415	.17	.05
2017	2nd	500	405	.19	.02
2018	3rd	500	390	.22	.03
2019	4th	500	378	.244	.024
2020	5th	500	365	.27	.026
2021	6th	500	353	.294	.024
			Imputed average marginal mobility for future calculations:		**.025**
School Year	Grade	Original Enrollment	Imputed Students still enrolled at a CPS school	Imputed Cumulative Mobility	Imputed Marginal Mobility
2022	7th	500	341	.319	.025
2023	8th	500	328	.343	.025
2024	9th	500	316	.368	.025
2025	10th	500	304	.393	.025
2026	11th	500	291	.417	.025
2027	12th	500	279	.442	.025

Appendix B: Timing of Cohorts

School year:	2014-15	2015-16	2016-17	2017-18	2018-19	2019-20*	2020-21	2021-22	2022-23	2023-24	2024-25
Cohort 1											
Intervention	CPC PreK	K	1st	2nd	3rd	4th	5th	6th			
Comparison	No PreK	K	1st	2nd	3rd	4th	5th	6th			
Cohort 2											
Intervention		CPC PreK	K	1st	2nd	3rd	4th	5th	6th		
Comparison		No PreK	K	1st	2nd	3rd	4th	5th	6th		
Cohort 3											
Intervention			CPC PreK	K	1st	2nd	3rd	4th	5th	6th	
Comparison			No PreK	K	1st	2nd	3rd	4th	5th	6th	
Cohort 4											
Intervention				CPC PreK	K	1st	2nd	3rd	4th	5th	6th
Comparison				No PreK	K	1st	2nd	3rd	4th	5th	6th

*SRI evaluation ends on Dec 1, 2020.

Appendix C: Propensity Score Matching/Weighting Process

Exhibit C-1 shows the characteristics of Cohort 1 children, characteristics of all kindergarten children enrolled in CPS in 2015–16 who met the eligibility criteria, and the propensity score weighted comparison group. The sample was diverse in terms of ethnicity, race, and family socioeconomic position.

To ensure that the method was successful at creating covariate balance, we compared the standardized mean difference between Cohort 1 CPC intervention children and their no-pre-K peers before and after propensity score weighting for each covariate. The standardized mean difference is the difference in means between the groups, divided by the pooled standard deviation of both the intervention and comparison groups. Standardized mean differences that are less than or equal to .25 indicate that the propensity score weighting successfully produced equivalent groups on a baseline characteristics or covariate (What Works Clearinghouse, 2014). The balance of the covariates in the analysis was greatly improved after applying the propensity score weighting method. Exhibit C-1 shows the balance before propensity score weighting, where the differences on covariates ranged from -0.32 to 0.55 standard deviations. After propensity score weighting (Column “Post-PSW Balance” in Exhibit C-1), the average differences on covariates ranged from -0.08 to 0.07 standard deviations. For example, before propensity score weighting, CPC children were less likely to be African American (-0.32), more likely to be Hispanic (0.46), more likely to speak Spanish at home (0.55), more likely to be from neighborhoods with a higher concentration of Hispanics (0.47) and a lower concentration of African Americans (-0.28), and more likely to come from neighborhoods with higher murder (0.54) and robbery (0.53) rates than their peers who did not enroll in CPC. After propensity score weighting, differences between CPC children and the comparison children were reduced to -0.08 to 0.07 (Column “Post-PSW Balance”) standard deviations on all baseline covariates. This analysis shows that the propensity score weighting did indeed create two well-matched groups of children.

Exhibit C-1. Intervention and Comparison Balance Statistics on Covariates Before and After Propensity Score Weighting

	Intervention			Weighted Comparison			Post-PSW Balance
Covariates	*N*	Mean	*SD*	*N*	Mean	*SD*	
Child characteristics							
Gender[a]	297	.47	.50	9445	.48	.09	-0.02
Black[a]	297	.30	.46	9445	.30	.08	-0.02
Hispanic[a]	297	.68	.47	9445	.67	.08	0.03
Speak Spanish at home[a]	297	.59	.49	9445	.58	.09	0.02
Neighborhood characteristics							
Percent Black[b]	297	.29	.42	9445	.30	.06	-0.03
Percent Hispanic[b]	297	.51	.38	9445	.50	.06	0.07
Household mean income (in thousands of dollars)[b]	297	53.9	18.4	9445	54.1	4.2	-0.05
Percent below poverty[b]	297	26.9	13.8	9445	27.1	2.3	-0.04
Percent low birthweight[c]	297	10.9	4.0	9445	11.0	0.7	-0.08
Counts of Murder[d]	297	30.7	12.2	9445	30.7	2.1	0.00
Counts of Robbery[d]	297	863.2	267.7	9445	860.5	46.9	0.04

Note: PSW= propensity score weighting.
[a] Child level data obtained from the school district for the 2015-2016 school year
[b] Census tract level variables are from 2013, the latest data available at the time of analysis (http://factfinder.census.gov/faces/nav/jsf/pages/download_center.xhtml). We assigned census tract based on the address of the school each child attended in kindergarten.
[c] Community area public health data are from 2009, the latest data available at the time of analysis (https://data.cityofchicago.org/Health-Human-Services/). We assigned community area based on the home zip code for children when they were in kindergarten (2015-16).
[d] Police district crime report data are from 2010, the latest data available at the time of analysis (http://home.chicagopolice.org/wp-content/uploads/2014/12/2010-Annual-Report.pdf). We assigned police district based on the home zip code for children when they were in kindergarten (2015-16).