Washington State Institute for Public Policy
110 Fifth Avenue SE, Suite 214 ● PO Box 40999 ● Olympia, WA 98504 ● 360.664.9800 ● www.wsipp.wa.gov

December 2020

Early Achievers Evaluation Report Two: *Pre-Kindergarten Quality and Child Outcomes in Kindergarten*

Revised February 2, 2021 for technical corrections

The 2015 Washington State Legislature passed the Early Start Act, which made participation in the state's quality rating and improvement system (QRIS), Early Achievers, mandatory for all early care and education programs serving non-school age children and receiving state funds.

In the Early Start Act of 2015, the legislature also directed the Washington State Institute for Public Policy (WSIPP) to evaluate the relationship between Early Achievers quality ratings and longitudinal outcomes for children who participate in subsidized child care and early education.[1] WSIPP was required to produce reports for the legislature in December 2019, 2020, 2021, and 2022. The final report is to include a benefit-cost analysis.

This second report provides a brief review of Early Achievers implementation, details our evaluation methodology, and defines the site and child study samples for WSIPP's evaluation. Our focus in this report is an evaluation of the impacts of Early Achievers during children's pre-k experience on their outcomes in kindergarten, including kindergarten readiness as assessed by the Washington Kindergarten Inventory of Developmental Skills (WaKIDS).

Summary

The 2015 Washington State Legislature passed the Early Start Act, which required all licensed child care facilities and early learning programs receiving state funds to participate in Early Achievers, the state's quality rating and improvement system (QRIS).

The Washington State Institute for Public Policy was directed to evaluate the impact of Early Achievers (EA) on long-term child outcomes and to produce a corresponding benefit-cost analysis.

In this second report we examine EA site status in children's final pre-kindergarten year as a predictor of children's kindergarten readiness. We present analyses for children in the Early Childhood Education and Assistance Program (ECEAP) and those receiving subsidy to attend a child care center.

Broadly, our findings do not reveal conclusive evidence that enrollment in an EA rated site significantly predicts kindergarten readiness. However, we do find that enrollment in a pre-kindergarten site with a rating at (or above) quality is a significant positive predictor of greater kindergarten readiness for children in ECEAP sites and for those with childcare subsidy.

Suggested citation: Goodvin, R., Rashid, A., & He, L. (2020). *Early Achievers evaluation report two: Pre-kindergarten quality and child outcomes in kindergarten* (Document Number 20-12-2203). Olympia: Washington State Institute for Public Policy.

[1] Second Engrossed Second Substitute House Bill 1491, Chapter 7, Laws of 2015.

Section I introduces WSIPP's assignment to evaluate Early Achievers, provides an overview of our planned report series, and outlines the research questions addressed in the present report. Section II briefly reviews QRIS background information and information on the implementation of the Early Achievers QRIS in Washington State. In Section III we provide an overview of data and define and describe the site and child study samples used in this evaluation. Section IV comprises details of WSIPP's data, research design, and results. Sections V details the limitations of the evaluation, summary, and conclusions.

Legislative Assignment

The Washington state institute for public policy shall conduct a longitudinal analysis examining relationships between the early achievers program quality ratings levels and outcomes for children participating in subsidized early care and education programs. (b) The institute shall submit the first report to the appropriate committees of the legislature and the early learning advisory council by December 31, 2019. The institute shall submit subsequent reports annually to the appropriate committees of the legislature and the early learning advisory council by December 31st, with the final report due December 31, 2022. The final report shall include a cost-benefit analysis.

2E2SHB 1491, Early Start Act of 2015

I. Introduction

The Early Start Act of 2015 (ESA) directed WSIPP to produce an evaluation of Early Achievers that addresses the relationship of quality ratings to child outcomes. The legislation specifies that WSIPP should assess children's outcomes over time in a longitudinal study. Additionally, the assignment specifies that WSIPP should assess outcomes for "children participating in subsidized early care and education programs." For WSIPP's evaluation, we understand this to include children in the state's Early Childhood Education and Assistance Program (ECEAP) as well as children supported by child care subsidies. See Exhibits 1 and 2 of WSIPP's Early Achievers Report One for a more detailed summary of subsidized child care and early learning (CC/EL) programs available to low-income children in Washington State and encompassed under this direction.[2] Finally, the legislature directed WSIPP to include a benefit-cost analysis in the final report.

Research Questions

The central task of this evaluation is to estimate associations of Early Achievers quality ratings and child outcomes. We frame this task in terms of the following three research questions, moving from broad to specific levels of estimation:

1) Does site participation in the QRIS process predict child outcomes?
2) Do differences in quality captured by QRIS ratings predict child outcomes?
3) Which QRIS standard areas (if any) best predict child outcomes?

In this report, we address the first two research questions specifically regarding Early Achievers during children's final pre-kindergarten year and their outcomes during kindergarten. We will address the third research question, regarding QRIS standard areas, in our 2021 report to the extent possible given data limitations. Research questions one and two, along with corresponding methods, are described in greater detail in Section IV of this report.

Early Achievers Evaluation Report Series

The Early Start Act of 2015 directed WSIPP to produce a series of four reports on the Early Achievers evaluation. This report is the second in the series. In this report, we focus on Early Achievers in the pre-k year and outcomes in kindergarten.

[2] Goodvin, R., & Hansen, J. (2019). *Early Achievers evaluation report one: Background and research design* (Doc. No. 19-12-2202). Olympia: Washington State Institute for Public Policy.

WSIPP's plan for the Early Achievers report series was based on several considerations, including the timeframe of Early Achievers implementation, the roll-out of the Washington Kindergarten Inventory of Developing Skills (WaKIDS) (a key outcome measure), and availability of mature child outcomes data. See Exhibit 1 for an outline of the planned report series.

For report three, due in December 2021, third-grade assessment data will not yet be fully mature. Our plan for report three includes addressing research question three (regarding quality standard areas) and conducting analyses on topics of special interest to the extent possible given data constraints.

This third report may also allow WSIPP to respond to emergent questions. Report four, due in December 2022, will focus on the impact of Early Achievers in the pre-k year on outcomes through the 3rd grade and include a benefit-cost analysis. The cohorts and data represented in the present report were largely unimpacted by COVID-19 related school closures. However, our plans for future reports may require adjustment to accommodate the impacts of these closures on Early Achievers, child care and early learning programs, and K-3 schooling and assessments. We discuss this in more detail in Section V.

Exhibit 1

Early Achievers Evaluation Report Series Plan

Report one: Dec 2019	Report two: Dec 2020	Report three: Dec 2021	Report four: Dec 2022
Background and research design	**Pre-k impact on kindergarten outcomes**	**Special topics***	**Pre-k impact on 3rd-grade outcomes; Benefit-cost analysis**
Describe Early Achievers implementation, review national evidence on QRIS in relation to child outcomes, summarize ratings progress to date, and outline planned research design as well as limitations of this evaluation.	**Impact of Early Achievers in the year prior to attending kindergarten (pre-k year) on child outcomes in kindergarten.** ***Address guiding questions one and two; address variation in effects.***	*Address guiding question three.* *Ex. Relationship of infant-early childhood quality ratings to pre-k and kindergarten outcomes.* *Ex. Within provider analysis of re-rates, renewal ratings, and child outcomes*	Impact of Early Achievers in the year prior to attending kindergarten (pre-k year) on child outcomes in 3rd grade. *Address three guiding questions; address variation in effects.* Benefit-cost analysis.

Note:

* WSIPP has identified a range of potential special topics that could be addressed in depth in report three. For each of these questions, and any additional questions that emerge, we must first determine whether available data support valid analyses.

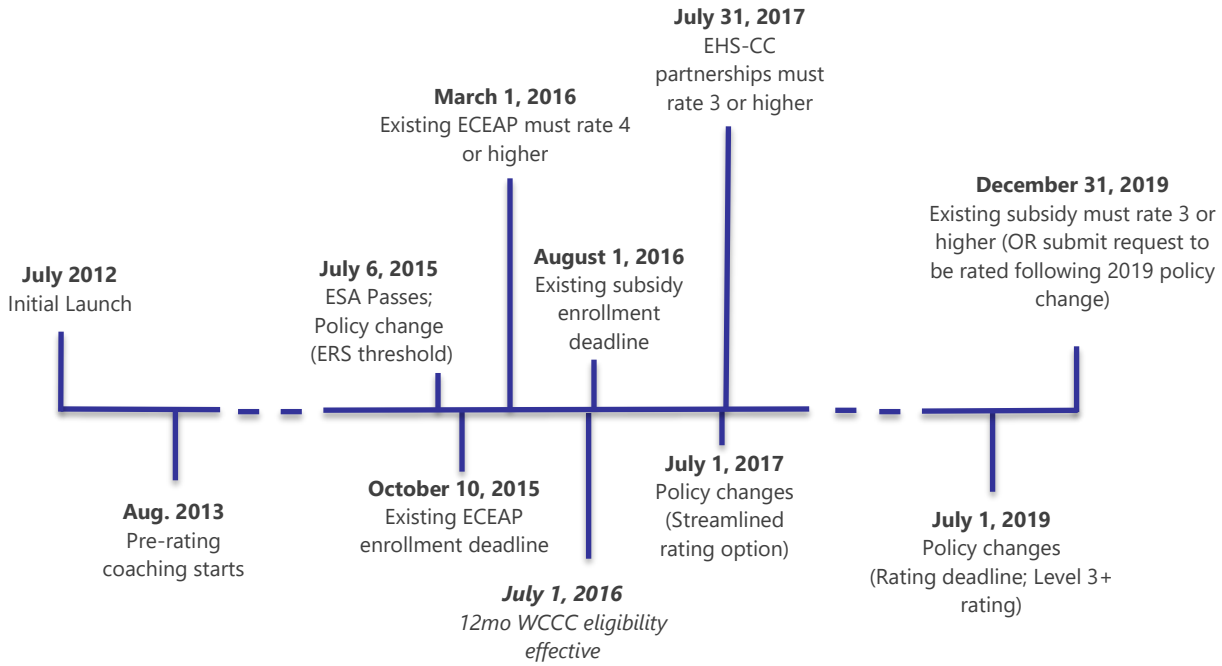
II. Background: QRIS and Early Achievers Implementation

In WSIPP's Early Achievers (EA) Report One we review definitions of child care quality and discuss the emergence of quality rating and improvement systems (QRIS) in the United States.[3] We also summarize the literature on child care quality and QRIS ratings as predictors of child outcomes. Briefly, evidence to date indicates that, on average, QRIS ratings are largely unrelated to child outcomes. Where associations have emerged, they have not been consistent across evaluations or child outcome domains. However, these research findings should be interpreted with caution given a range of methodological limitations.

In EA Report One Section IV we also review the implementation of Early Achievers in Washington State.[4] In the following section, we briefly highlight several key points regarding implementation. Early Achievers initially rolled-out by region from July 2012 through July 2013. Passage of the Early Start Act in July 2015 made Early Achievers mandatory for sites serving non-school age children with state funding. The Early Start Act (ESA) also set timelines for participation for ECEAP and non-ECEAP sites serving children on subsidy (Exhibit 2).

Exhibit 2

Early Achievers Implementation Timeline as of December 2019



Notes:
Existing ECEAP: The site had contracted ECEAP slots as of July 1, 2015.
Existing subsidy: The site received a subsidy payment between July 1, 2015, and June 30, 2016.

[3] Goodvin & Hansen (2019).

[4] Ibid, Sec. IV.

Early Achievers Overview

Early Achievers participation is open to all licensed or certified child care and early learning sites in Washington State. Sites serving children with state funding, either through ECEAP or through child care subsidy, are required to participate.

Prior to receiving an initial quality rating sites may engage in supports that include coaching and consultation. Additionally, some may receive need-based grants or professional development scholarships.

Sites request an on-site evaluation, and subsequently, receive a rating from Level 2 to Level 5 based on points earned across the Early Achievers quality standard areas.[5]

Sites earn points in five quality standard areas (See Exhibit 3):

1) Learning environment and interactions;
2) Child outcomes;
3) Curriculum and staff support;
4) Professional development and training; and
5) Family engagement and partnership.

The rating process and rating requirements differ for ECEAP and non-ECEAP sites serving children on subsidy. ECEAP sites must be rated at a Level 4, and non-ECEAP subsidy sites must achieve a Level 3 to satisfy requirements and be considered at-quality.[6]

Sites that do not meet this expectation engage in additional quality improvement supports and are re-rated, typically within approximately one year. All sites are expected to renew their Early Achievers rating on a three-year cycle.

Exhibit 3

Overview of Early Achievers Points by Level

Note:
Source: Adapted from Early Achievers Participant Operating Guidelines, January 2020.

Recent Changes to Early Achievers

Several changes to Early Achievers were enacted in 2019 and 2020.[7] However, our evaluation will only include ratings assigned through the 2019-2020 academic year, given the time required for outcome data maturation as well as our reporting timeline. As a result, WSIPP's data and analyses will not reflect these changes.

The exception is a 2019 policy change directing DCYF to introduce a Level 3+ rating. This change differentiates sites with a higher number of points within the large group of sites initially rated a Level 3. WSIPP will apply the Level 3+ rating retroactively.

[5] DCYF's Early Achievers Operating Guidelines comprehensively describe the system. Early Achievers Participant Operating Guidelines, January 2020.
[6] Engrossed Second Substitute House Bill 1391, Chapter 369, Laws of 2019.
[7] Beginning in July 2019, Early Achievers phased out the initial "Level 2" professional development trainings and introduced an optional baseline observational assessment using the ERS. Additionally, Early Achievers anticipated shifting in late 2020 from using two observational assessments (ERS and CLASS) to using a single observational assessment (ERS-3).

III. Outcome Evaluation: Data and Sample Construction

In this section, we describe data sources and key information required for WSIPP's evaluation. We also describe our target population and outline our approach to constructing our evaluation sample.

Data Sources

For all analyses, we use state administrative records. Data come from the Department of Children, Youth, and Families (DCYF); the Office of the Superintendent of Public Instruction (OSPI); and the Department of Health (DOH). The Education Research and Data Center (ERDC)[8] provided integrated identity matching for all child-level records, described later in this section. See Appendix I for more detail on data sources.

Estimating impacts of Early Achievers on child outcomes requires, at minimum, records of CC/EL enrollments for sample identification, Early Achievers participation and ratings for the CC/EL sites children attended, and relevant child outcome data.

Beyond CC/EL sites' Early Achievers participation and ratings, we must also account for other potential sources of site-, school-, community-, family-, and child-level variation in children's CC/EL experiences and long-term outcomes. Therefore, we incorporated ECEAP and child care subsidy site characteristics from DCYF administrative data, Early Achievers coaching data from DCYF/Child Care Aware, child and family characteristics from birth records, CC/EL sites' neighborhood characteristics from census tract records, and school characteristics from OPSI public records.

Key Data

Child Enrollments

DCYF provided ECEAP enrollment and child care subsidy records to WSIPP through a data partnership with ERDC.

Child ECEAP participation in a given academic year (AY) is indicated by an ECEAP enrollment record for that year. Participation in subsidized child care in a given AY is indicated by one or more monthly subsidy payment records during that year. Both ECEAP enrollment and child care subsidy payment records indicate the specific site(s) where the child was enrolled, allowing WSIPP to identify the Early Achievers status and quality level of the site that the child attended during their pre-kindergarten year.

[8] For additional information on the ERDC, please see ERDC's website.

Pre-Kindergarten Early Achievers Ratings

DCYF/Cultivate Learning provided records on ECEAP and subsidy sites' participation in Early Achievers. For each site in our sample, for each AY in our study period, we identify the following:

- Rating status (rated vs. unrated) and
- Rating level at-quality.[9]

Child Outcomes

The ERDC maintains a longitudinal data system to facilitate linking child-level records across ECEAP, child care, and K–12 schools data owned by DCYF and OSPI, respectively. ERDC matched children who participated in ECEAP or subsidized child care during our study period to kindergarten enrollment and outcomes data.

ERDC provided WSIPP with records for the following outcomes during children's kindergarten enrollment:

- Kindergarten readiness assessments,
- Special education placement,[10]
- Absences,
- Exclusionary discipline, and
- Grade retention (enrollments).[11]

Additionally, ERDC provided all available ECEAP kindergarten readiness assessment records for our study period.

Kindergarten Readiness. The primary outcome of interest in this report is children's kindergarten readiness, observed within the first two months of entering kindergarten. Kindergarten readiness is assessed using the Washington Kindergarten Inventory of Developing Skills (WaKIDS), which documents teachers' observations of children's knowledge, skills, and abilities.[12]

Additionally, children in ECEAP are assessed each Fall and Spring using the Teaching Strategies GOLD (TSG), an assessment that serves as the foundation for the WaKIDS.[13] We also examine kindergarten readiness for ECEAP students in the spring of their pre-kindergarten year.

On both the WaKIDS and TSG, teachers observe child skills in six domains: social-emotional, physical, cognitive, language, literacy, and mathematics. Children are considered "kindergarten ready" in a given domain if they meet or exceed a benchmark score indicating age-appropriate skills. Children who meet or exceed the benchmark in all six domains are identified as being "kindergarten ready."[14] In this report, we predict child readiness on five of six, and six of six domains.[15]

[9] ECEAP sites must rate at a Level 4 to be considered at-quality, and subsidy sites must rate at a Level 3. Although ratings range from Level 2 through Level 5, currently few sites have received ratings higher than what is required to be considered at-quality (See Exhibits 8 and 9). WSIPP is thus unable to complete analyses using the full possible range of scores.

[10] Defined by the presence of at least one enrollment in the CEDARS special education programs data file.

[11] Exclusionary discipline and grade retention occurred too rarely in our kindergarten sample to complete analyses. We will review these outcomes again for third grade analyses but do not address them further in this report.

[12] WaKIDS was legislatively mandated to be part of state-funded full-day kindergarten in the 2012-13 school year (RCW 28A.150.315 and RCW 28A.655.080). All Washington schools were reporting WaKIDS data starting in 2017-18.

[13] The WaKIDS assessment utilizes a custom subset of items from the Teaching Strategies GOLD assessment that is used by ECEAP sites.

[14] Education Research & Data Center. (2018). *Early learning feedback report.*

[15] Several changes to the Teaching Strategies GOLD assessment and to the WaKIDS item subset were implemented during our study period, with implications for

Target Population and Cohorts

We selected target population cohorts based on the timing of Early Achievers implementation. Additionally, we considered the timing of data maturation and availability, as well as the timing of the WaKIDS roll-out across schools.

We define cohorts using academic years (AY) running from September through August. This AY approach aligns with standard birthdate cutoffs for both ECEAP and kindergarten entry, the ECEAP academic year program, and K-3 outcome data. The target population for research questions addressed in this report is children who attended ECEAP or subsidized child care in their pre-kindergarten year in AY 2014-15 through AY 2017-18 and had their first enrollment in kindergarten in the following year.

Exhibit 4 illustrates the extent of Early Achievers adoption for both ECEAP and subsidy sites for our target population Cohorts 1 through 4. We note the expected pre-kindergarten, kindergarten, and 3rd-grade year for each cohort.

In Exhibit 4 we also note the percentage of elementary schools implementing WaKIDS for each kindergarten cohort. One implication of WaKIDS adoption over the same period as the initial Early Achievers roll-out is that through AY 2016-17 the WaKIDS was not conducted for all children impacted by Early Achievers.[16] In Cohort 1 of our analysis sample (see Exhibit 4), approximately 70% of schools were administering the WaKIDS (71% in our ECEAP sample and 76% in our non-ECEAP subsidy sample). For Cohort 2, over 99% of schools were administering the WaKIDS.

Pre-K to K Outcomes

Children's experience with an Early Achievers site during their pre-kindergarten enrollment is the focus of this report. We examine the impact of pre-kindergarten quality experience on kindergarten outcomes for four cohorts of children.[17]

this evaluation. See Appendix IV for detail on these changes and on our approach to mitigating their impact.

[16] The WaKIDS was initially rolled out as a requirement tied to funding for full-day kindergarten, starting with schools reporting the highest rates of children qualifying for free and reduced-price meals (FRM).

[17] We received data for two additional cohorts but were unable to include them in this analysis. Children attending pre-kindergarten in AY 13-14 and kindergarten in 14-15, were assessed using an earlier version of the WaKIDS (see Appendix IV for additional information). For children attending pre-kindergarten in AY 18-19 we were missing too many covariates to produce analyses comparable to those for other years. We tested the robustness of our main analyses to inclusion of the AY 18-19 pre-k cohort and omission of missing covariates (see Appendix VI).

Exhibit 4

Early Achievers Roll-Out by Academic Year Cohorts in Target Population

Cohort	Birth year	1- to 2-year-old	2- to 3-year-old	3-year-old pre-k year	**4-year-old pre-k year**	Expected K year (WaKIDS)	Expected 3rd-grade year	% of schools using WaKIDS
ECEAP sites								
1	9/09 - 8/10	2011-12	2012-13	2013-14	**2014-15**	2015-16	2018-19	75%
2	9/10 - 8/11	2012-13	2013-14	2014-15	**2015-16**	2016-17	2019-20*	95%
3	9/11 - 8/12	2013-14	2014-15	2015-16	**2016-17**	2017-18	2020-21*	100%
4	9/12 - 8/13	2014-15	2015-16	2016-17	**2017-18**	2018-19	2021-22	100%
Subsidy sites								
1	9/09 - 8/10	2011-12	2012-13	2013-14	**2014-15**	2015-16	2018-19	75%
2	9/10 - 8/11	2012-13	2013-14	2014-15	**2015-16**	2016-17	2019-20*	95%
3	9/11 - 8/12	2013-14	2014-15	2015-16	**2016-17**	2017-18	2020-21*	100%
4	9/12 - 8/13	2014-15	2015-16	2016-17	**2017-18**	2018-19	2021-22	100%

Notes:

Before EA
Limited, voluntary EA ratings
First required EA ratings
EA ratings approximately 50% complete
EA ratings 70-100% complete

* Cohort will have some degree of disruption due to COVID-19 school closures, which may affect WSIPP's approach to evaluation in 2021 and 2022 reports.

Sample Construction

Exhibit 5 depicts an overview of the steps we took to construct ECEAP and subsidy samples for the analyses presented in this report. All steps were completed separately for ECEAP and subsidy site records. [18]

Foundational Samples

Initial steps in this process bring together site- and child-level data sources to construct a complete observed sample. This sample provides a foundation for further specification of the site and child samples most appropriate to address individual research questions.

We first integrated site-level records from multiple sources over time for all ECEAP and non-ECEAP licensed sites that participated in Early Achievers.

Second, we used child enrollment records to identify all children in ECEAP or subsidized child care during our target cohort years.

Third, we linked child participation records to each site attended, for each year of child attendance. In this linked file we observe each year and site of a child's CC/EL enrollment and the site's corresponding Early Achievers rating status and quality rating during that year.[19]

Analysis Samples

All additional steps in sample construction were taken to define analysis samples used for the specific research questions addressed in the present report.

Specifically, in the next step, we restricted our sample to children who were observed in pre-kindergarten (ECEAP or subsidized child care) and also in kindergarten enrollment data in the following AY.[20] Next, we selected children's first enrollment site during the AY prior to their first kindergarten enrollment and refer to this site as the child's pre-kindergarten (pre-k) enrollment.[21]

Following this step, we applied additional restrictions to sites and children included in the final study samples as appropriate to our specific research questions, methods, and data. All sample construction steps, including these final sample restrictions, are discussed in more detail in Appendix II.

[18] We refer to these groups as ECEAP sites and subsidy sites, respectively. However, we note that under Washington's mixed CC/EL delivery system there is overlap between ECEAP and subsidy sites, and many of these sites provide additional services. Separate samples and analyses accommodate differences in the two types of CC/EL settings, and differences in Early Achievers participation timelines and requirements.

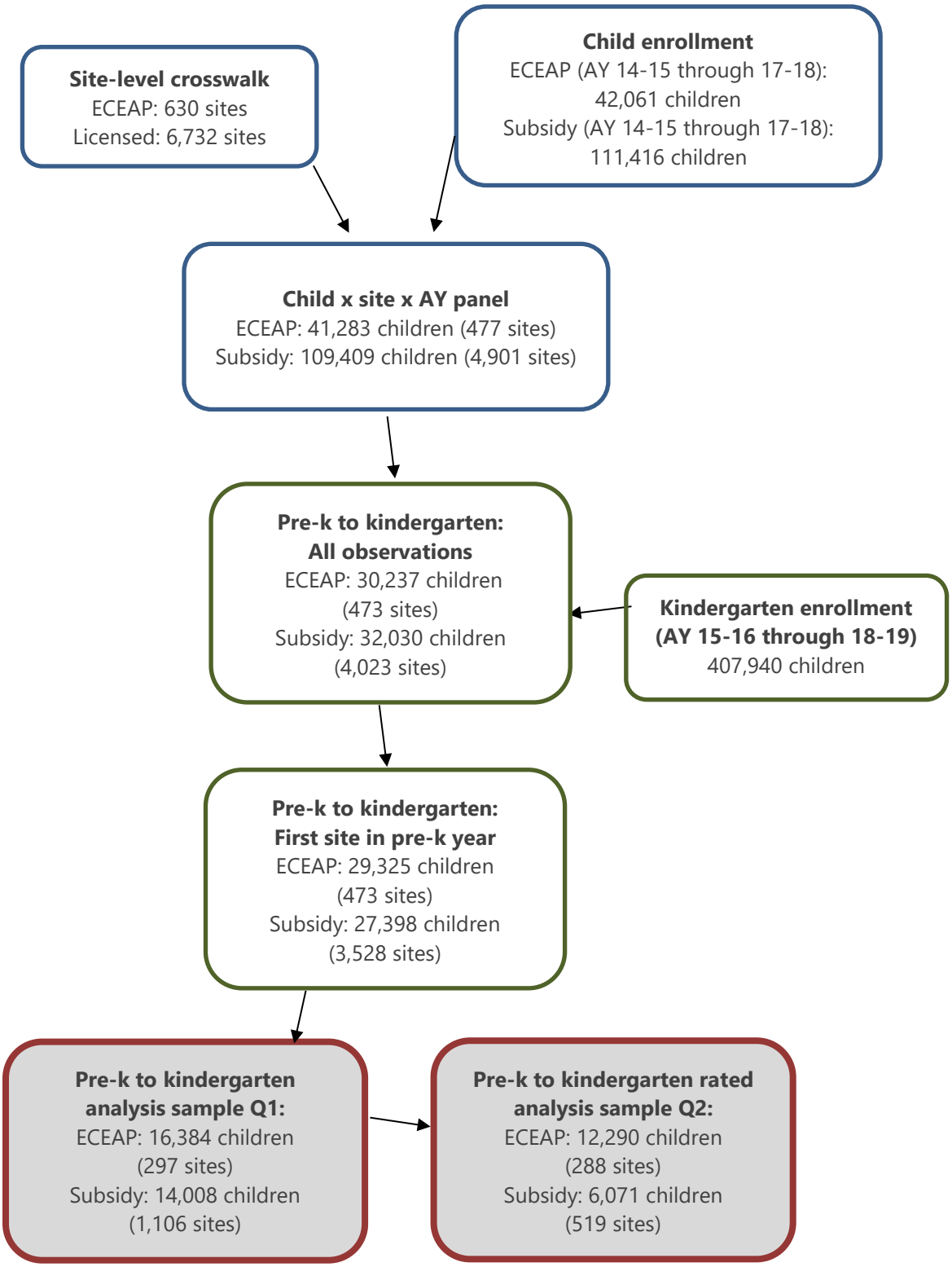
[19] Imperfect historic record systems result in loss of both sites and children in this step. Specifically, not all sites observed in the site-level records had observed ECEAP or subsidy child enrollments, and not all children in enrollment records could be accurately matched to a site given available site identifiers. Some non-ECEAP licensed sites may not have served children on subsidy.

[20] In this step we selected children observed in kindergarten enrollment records in our target years and in the year prior to their first kindergarten enrollment, without regard to year of expected kindergarten enrollment based on birth cohort. Of children observed in a pre-k setting at age four, about 90% of the ECEAP sample, and about 72% of the subsidy sample, was also observed in kindergarten. Kindergarten enrollment outside of the "expected" year was rare in our sample. For ECEAP 0.2% of children appeared to be enrolled in kindergarten one year earlier than expected based on birth cohort. For subsidy 0.5% appeared to be enrolled one year earlier than expected, and 2% appeared to be enrolled one year later than expected based on birth cohort.

[21] Children may have multiple enrollments in their pre-kindergarten year, both simultaneous and consecutive. We excluded children whose first pre-k enrollment occurred in more than one site simultaneously. This exclusion is reflected in the sample sizes in Exhibit 4 corresponding to "first site in pre-k year."

Exhibit 5

Overview of Early Achievers Sample Construction



III. Outcome Evaluation: Sample Description

In this section, we describe WSIPP's ECEAP and child care subsidy analysis samples. Here, and throughout this report, we present information separately for children in ECEAP and subsidy sites.

WSIPP's main analysis samples for this report includes children in ECEAP or subsidized child care in the 2014-15 AY through 2017-18 AY.

Early Achievers Ratings

As noted in Exhibit 3, sites started receiving Early Achievers ratings in July 2013. When the Early Start Act passed in July 2015, ECEAP and subsidy sites were already participating in Early Achievers. Exhibits 6 and 7 show, for our analysis sample, the probability over time that pre-kindergarten enrollments were in a rated site.

These figures are consistent with growth in the number of rated sites across our study period for both ECEAP and subsidy. Consistent with the mandated rating timeline for existing ECEAP sites, by the 2015-16 AY over 80% of ECEAP pre-kindergarten enrollments in our sample are in rated sites (Exhibit 6). We also observe that pre-k enrollments in a rated subsidy site increase relatively more steadily across our study period (Exhibit 7).

Exhibit 6

Probability of ECEAP Pre-K Enrollment in a Rated Site

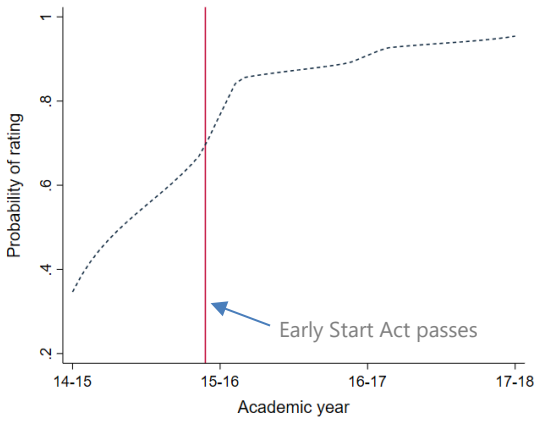


Exhibit 7

Probability of Subsidy Pre-K Enrollment in a Rated Site

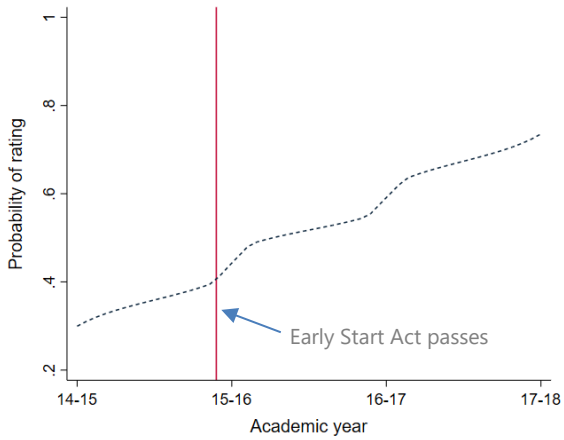
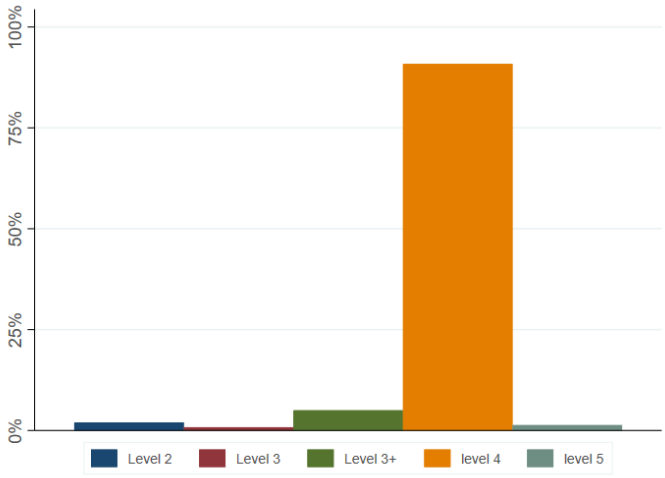


Exhibit 8

Distribution of ECEAP Pre-K Enrollments in Sites Rated 2 through 5



Exhibits 8 and 9 show, for our ECEAP and subsidy analysis samples, respectively, the distribution of pre-k enrollments across Early Achievers quality rating levels.

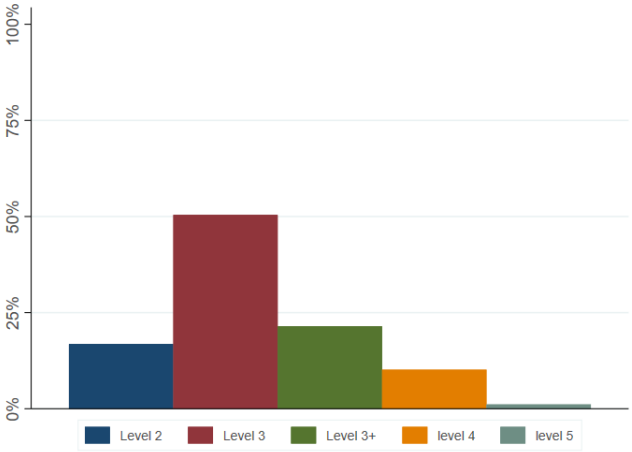
As seen in Exhibit 8, nearly 90% of ECEAP pre-k enrollments in our sample are at sites rated Level 4. This is consistent with the rating requirement for ECEAP sites.

For subsidy sites, a Level 3 quality rating is considered at-quality. Approximately 70% of pre-k enrollments observed in our sample are at sites rated Level 3 (Exhibit 9).[22]

In sum, for both ECEAP and subsidy sites variation in overall quality ratings is extremely limited. This restricted rating level variation across sites limits our ability to confidently estimate the relationship between rating level and child outcomes. Implications of this for WSIPP's analyses are discussed in the following sections.

Exhibit 9

Distribution of Subsidy Pre-K Enrollments in Sites Rated 2 through 5



[22] This sample distribution is consistent with but not identical to the distribution in the population. We address correspondence between sample and population in greater detail in Appendix II and III.

Child Outcomes

As seen in Exhibit 10, 61% of children in our sample of ECEAP sites meet/exceed the benchmark on at least 5 of 6 WaKIDS domains, and 39% meet/exceed the benchmark on all 6 domains. Children in subsidy sites show a slightly lower average with about 53% meeting the benchmark on at least 5 domains and 35% meeting the benchmark on all 6 (Exhibit 11).[23]

An additional outcome of interest is special education enrollment in kindergarten. This is a relatively rare outcome, observed in 12% of children who attended an ECEAP site and 9% of children who attended a subsidy site. Finally, we observe the average number of unexcused monthly absences at roughly 1.22 and 1.08 for ECEAP and subsidy attendees, respectively (Exhibits 10 and 11).

Exhibit 10

Child Outcomes for ECEAP Sample

Outcome	All sites (1)	Rated sites (2)
Number of domains meet/exceed WaKIDS:		
At least 5	0.61	0.61
	(0.49)	(0.49)
All 6	0.39	0.39
	(0.49)	(0.49)
Number of domains meet/exceed TS Gold:		
At least 5	0.81	0.81
	(0.39)	(0.40)
All 6	0.60	0.59
	(0.49)	(0.49)
Special education enrollment	0.12	0.11
	(0.32)	(0.32)
Average monthly unexcused absences	1.22	1.22
	(1.08)	(1.07)
Observations	10,278	7,844

Note:
Cells report the unweighted sample mean with the corresponding standard deviation in parenthases underneath.

Outcome averages are similar for the sample of children attending ECEAP or subsidy sites that have already been rated (Exhibits 10 and 11, Column 2).

[23] Although meeting kindergarten readiness benchmark scores on all domains is considered optimal, in our target population the rates of kindergarten readiness across all six domains are relatively low. Consistent with our sample, the ERDC reports that between 30% and 37% of lower income and ECEAP students are kindergarten ready on all six domains during our study period. (ERDC. (2018). *Early learning feedback report.)*. Given these rates, readiness on all six domains may be too high a threshold by which to detect impacts of CC/EL interventions such as Early Achievers. In other words, we may observe program impacts for readiness on five of six domains, even if six of six remains elusive.

Site and Child Sample Description

ECEAP and subsidy sites in our analysis samples represent all Child Care Aware regions. We observe enrollments before or during the 14-15 AY for most sites in our sample (87% for ECEAP and 97% for subsidy), indicating that our sample consists largely of sites that existed prior to the passage of the Early Start Act. Sites in our samples are almost entirely center-based.[24]

Most children in our ECEAP analysis sample attend a part-day program, while most children in the subsidy analysis sample have full-day child care subsidies. The average age for pre-k enrollment is nine months for the ECEAP sample, and eight months for the subsidy sample.[25] Children in our sample are mostly white (ECEAP—35%, subsidy—43%) or Hispanic/Latinx (ECEAP—45%, subsidy—31%). English is identified as the primary language for most children in both samples. Approximately 50% of both samples are female.[26]

Exhibit 11

Child Outcomes for Subsidy Sample

Outcome	All sites (1)	Rated sites (2)
Number of domains meet/exceed WaKIDS:		
At least 5	0.53	0.54
	(0.50)	(0.50)
All 6	0.35	0.35
	(0.48)	(0.48)
Special education enrollment	0.09	0.09
	(0.28)	(0.28)
Average monthly unexcused absences	1.08	1.09
	(1.02)	(1.02)
Observations	9,153	4,380

Note:
Cells report the unweighted sample mean with the corresponding standard deviation in parenthases underneath.

[24] In ECEAP records, we observe very few family home child care (FHCC) sites—only about 1.3%. In our records, approximately 66% of children on subsidy attend a child care sites classified in licensing data as FHCC. However, when we restrict our analysis sample to children observed in both kindergarten and the year prior, only 25% of sample enrollments are in FHCC suggesting that these sites disproportionately serve younger children. When we add additional sample restrictions, 10% of enrollments for our analysis sample are in FHCC, with only 1% in FHCC having received a rating. We include additional comparison of site-level population characteristics to our analysis and rated analysis samples in Exhibits A6 and A7 of Appendix III.

[25] Duration is based on first and last enrollment date, and may thus overcount duration where enrollment was not continuous. Enrollment during the pre-k year ranged from 1 to 12 months. Results are robust to the exclusion of children enrolled for less than 6 months during the pre-k year.

[26] Descriptive statistics for characteristics of site- and child-level analysis samples are presented in the Appendix III.

IV. Method and Results

Question One: Does Site Participation in the QRIS Impact Child Outcomes?

One of the most important potential effects of Early Achievers is inducing providers to improve the quality of care offered to children. That quality improvement may be observed in subsequent ratings. However, by design, when sites receive a rating, they typically meet the required quality rating level (at least a Level 3 for subsidy sites and at least a Level 4 for ECEAP sites). Consequently, the most important effects of Early Achievers may not be captured by rating differences among facilities. Rather, sites' progression through Early Achievers may offer the clearest evidence of impact on child outcomes. We focus here on sites that have received a rating, relative to sites that have not yet achieved this milestone, to mark progression through Early Achievers.[27]

Method

In an ideal research setting, we would randomly assign children to rated or unrated CC/EL sites, or to sites of varying quality levels, to assess the unbiased relationship between the program quality rating and academic outcomes. However, in reality, parents choose which early learning site(s) their children will attend. Additionally, WSIPP's evaluation of Early Achievers is retrospective, and we are unable to use a controlled trial in which we randomly assign children to differing CC/EL experiences.

Therefore, an evaluation of the relationship between rating status or quality and child outcomes may suffer from self-selection of individuals. That is, it may be the case that children who would tend toward better outcomes regardless of CC/EL quality are most likely to attend sites that are already rated and/or receive a higher rating, such as families with greater resources or connections. Alternatively, it may be the case that children who would tend to have the least positive outcomes regardless of site quality are most likely to attend sites that are already rated and received a higher quality level. For example, children with greater cumulative risk may be given attendance priority in higher quality sites.[28] In the former situation, child-level selection may lead us to erroneously overestimate the positive impact of QRIS ratings on academic outcomes; the latter scenario may lead to erroneously under-estimating the impact. Notably, both may occur, leading to complex selection issues and possibly null or mixed findings.

To address the possible bias arising from the fact that parents choose which CC/EL to attend ("child-level selection"), we use a statistical approach known as entropy balancing. Under random assignment, we would expect no difference in characteristics between treatment and comparison group members, and entropy balancing aims to mimic this condition (see the sidebar on the following page).[29]

[27] Additional milestones, such as submission of a request for an on-site rating, may indicate earlier stages of this progression. Due to high rates of missing data for other milestone dates we focus exclusively on rating receipt.

[28] ECEAP sites actively prioritize enrollment of children with more risk factors (among eligible and allowable children), as described in the ECEAP performance standards. Non-ECEAP subsidy sites do not follow uniform eligibility or enrollment guidelines. Sites with ready access to child mental health supports or direct training in offering trauma-informed care may be more likely to accept children with greater social and emotional needs, as discussed in the 2016 Final Report of the Children's Mental Health Workgroup.

[29] Hainmueller, J. (2012). Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Political Analysis*, 25-46.

Program Evaluation with Entropy Balancing

Entropy balancing is a method that reweights the comparison observations (e.g., children who attend an unrated site) such that they are identical to treatment observations (e.g., children who attend a rated site) on the means and variance of observable characteristics.[#] In other words, this "matching method" ensures that in our outcomes analysis we are comparing children who are similar across observable traits such as familial characteristics and school environment.

In Appendix V (Exhibits A16 and A17), we demonstrate that entropy weights achieve improved balance across a rich set of child-level and kindergarten school-level characteristics—e.g., duration in subsidy/ECEAP care, sex, mothers' age and education at time of birth, and racial makeup of the school attended for kindergarten.[^]

Notes:
[#]Throughout the primary analysis, we balance covariates on both the mean and variance of select characteristics. However, for several subgroup and sensitivity analyses we force balance only on the mean of covariates due primarily to small sample size restrictions.
[^] More discussion regarding the implementation of entropy weighting, and its comparative advantage over regular multivariate regression analysis or alternative weighting and matching methods can be found in Section IV of the Appendix. The sensitivity of select baseline results to a nearest-neighbor Propensity Score Matching method can be found in Appendix VI, Exhibit A25.

Just as children are not randomly assigned to early learning sites, the sites themselves are not randomly assigned to participate in the QRIS program. Sites choose when to complete the rating assessment (by the deadline). Therefore, it is possible that sites with higher quality at baseline were more likely to register for Early Achievers (particularly prior to the ESA) and to complete ratings and are thus overrepresented among rated sites.

To relieve site-level concerns like the one mentioned above, we combine our entropy balancing method with a fixed-effects empirical strategy to address research question one (see the sidebar on the following page). Site fixed effects will account for all differences (observable and unobservable) between sites that predict child outcomes and do not change over time (e.g., the physical size of the site and director motivation). In addition, we include year fixed effects to account for year-to-year differences that predict outcomes and are shared by all sites (e.g., an economic recession).[30]

[30] A causal interpretation of our fixed-effects estimation strategy requires us to assume that outcomes for children in treated sites would have evolved similarly to those of children in untreated sites in the absence of treatment. Due to limitations in the data, we are not able to assess the validity of this assumption and as such caution against such causal interpretation. For more discussion on fixed effects estimation strategies and limitations see Appendix V and VII.

Program Evaluation with Fixed Effects

The **fixed effects** empirical strategy allows us to identify the impact of site rating by exploiting within-site variation in rating status instead of between-site variation. That is, we compare changes in child outcomes that occur before and after a site is rated, not differences in child outcomes between rated and unrated sites.

An advantage of the fixed effects strategy is that we do not need to have information on each and all possible (time-invariant) differences between sites. These differences are all accounted for at once with the specification of a site fixed effect.
Fixed effects notably do not control for differences between sites that change over time. Therefore, our models also include site-level control variables that are time-varying such as racial demographics, population, and economic indicators.[#]

We estimate our fixed effects model using Ordinary Least Squares regression analysis.[^] Standard errors are estimated to allow for clustering at the pre-k site-level.

For more detail on these and other methods used in analyses to address research question one, see Appendix V.

Notes:
[#]These controls include average monthly enrollment (subsidy), average annual enrollment (ECEAP), and the following census tract information: percent at least high school graduate, percent at least college graduate, percent at households below the poverty line, unemployment rate, median household income, population under the age of five, percent black, percent Hispanic/Latinx, and percent white.
[^]All outcomes analyses control for the aforementioned child-level, school-level, and site-level characteristics. Standard errors are estimated to adjust for clustering at the site-level.

Main Results

We summarize the results of our regression analysis using confidence interval plots. Our plots summarize the relationship between attending a rated early learning site and child outcomes—this relationship is estimated using within-site variation in rating status over time.

Kindergarten Readiness, Fall of Kindergarten. Figure A of Exhibit 12 shows how the kindergarten readiness of children who attend rated ECEAP sites compares against the kindergarten readiness of children who attend sites that have not yet been rated. We include two measures of readiness based on the WaKIDS assessment: meeting/exceeding benchmark scores in at least five domains and meeting/exceeding in all six domains.[31]

The first estimate (shown with a red square) plotted in Figure A of Exhibit 12 indicates that, on average, children who attend a rated ECEAP site are 6% more likely to be kindergarten ready based on meeting/exceeding benchmark standards in at least 5 domains (i.e., the "impact" of attending a rated site is 6%). About 61% of children meet/exceed in at least 5 domains. When looking at all 6 domains (shown with a blue diamond), estimates suggest that children who attend a rated ECEAP site are 10% more likely to meet/exceed all domains compared to children who attend a site that is not yet rated—about 40% of children in the sample meet all 6 domains.

[31] Detailed results are summarized in Appendix VI. The primary results tabulated in the Appendix include estimates corresponding to a third outcome: meet/exceed at least four domains.

As in all statistical analysis, there is some degree of uncertainty about the true impact of attending a site that has completed an Early Achievers rating. The points in the plot show our best estimates of the true program impacts, and the vertical bars around them show the 90% confidence interval corresponding to that estimate. In other words, we are 90% confident that the true relationship between attending a rated site and the outcome lies somewhere within that bar.

In Figure A of Exhibit 12, the confidence interval bars all cross the zero line which indicates that it is possible that the impact of rating is zero, and that in this context the program has no effect. However, the relatively long bars indicate that the estimates are imprecisely measured. Throughout this analysis we often face imprecise estimates as a result of our relatively small comparison sample—for example, only 20% of the sample attend an ECEAP site prior to rating completion.[32]

In sum, attending a rated ECEAP site is associated with modest improvement in kindergarten readiness, but results are ultimately inconclusive.

Unlike ECEAP sites, the results depicted in Figure B of Exhibit 12 suggest that there is no relationship between attending a rated subsidy site and WaKIDS performance—there is no convincing evidence that impacts are different from zero.[33]

Kindergarten Readiness, Spring of Pre-K. As discussed in Section III, children in ECEAP are also assessed for kindergarten readiness in the spring of their pre-k year. Parallel analysis indicates that attending a rated ECEAP site has no significant relationship with TS Gold assessment performance in the spring of a child's pre-k year. Results and further discussion regarding this analysis can be found in Appendix VI.

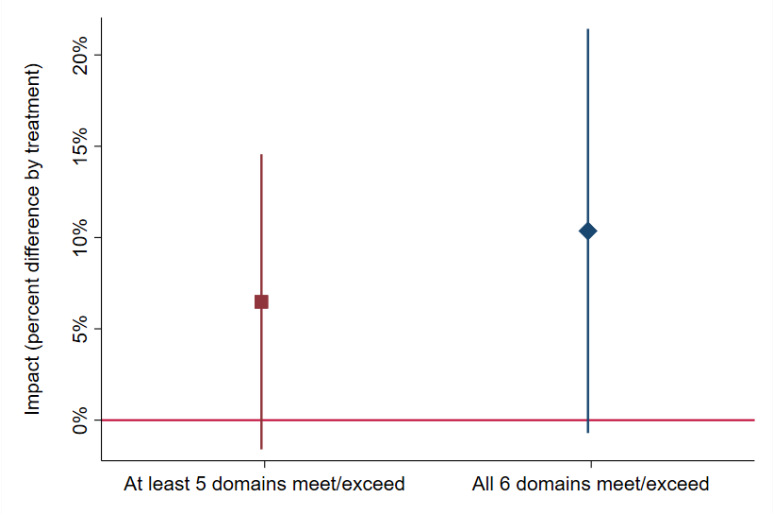
[32] Our treatment variable is an indicator that takes on a value of 1 it the child attends a rated site and 0 otherwise. The more variation in the distribution of treatment values the more precisely we can estimate the impact associated with treatment.

[33] We additionally explore differential associations across racial/ethnic subgroups. Although inconclusive, results indicate that any positive impact associated with attending a rated site are primarily driven by Hispanic/Latinx attendees for both ECEAP and subsidy sites. In addition, we explore differential effects across duration of childcare tenure (one year vs. one years plus). More in-depth analysis and discussion regarding the subgroup analyses can be found in Appendix VIII.

Exhibit 12

Pre-K Early Achievers Rated Status and Kindergarten Readiness

(A) ECEAP site rated status and WaKIDS domains ready



(B) Subsidy site rated status and WaKIDS domains ready

Impact: Percentage difference in outcome between the treated and comparison group. In Figure A of Exhibit 12 above, the estimated treatment impact on readiness in at least 5 domains is about 6%
Confidence interval: Range of values that likely includes the *true* treatment effect. In Figure A of Exhibit 12 above, the confidence intervals around the estimated treatment impact on readiness in at least 5 domains suggest that the true (population) impact lies between -2% and +15%. In other words, children attending a rated pre-k site could be as much as 2% less likely, *or* 15% more likely, to be kindergarten ready on at least 5 domains.

Notes:

Each point represents the impact of the marginal effect derived from a sperate regression model, and the lines represent the corresponding 90% confidence interval.

Each model controls for the full set of control variables and adjusts standard errors for clustering at the pre-k site level. Where WaKIDS outcomes are used the model additionally controls for the first year of implementation in school.

Additional Kindergarten Outcomes. Results shown in Exhibit 13 summarize the relationship between attending a rated site and additional outcomes measured in the kindergarten year—for ECEAP and subsidy sites, respectively. Additional outcomes include the probability of special education enrollment and the average number of monthly absences.

Estimates indicate that children who attend a rated ECEAP or subsidy site are roughly 15% less likely to be enrolled in a special education program (shown in red symbolized with a square)—the average special education enrollment rate for ECEAP site attendees is 11% and 8% for subsidy enrollees. The estimated relationship between site rating and special education enrollment is statistically insignificant.

In Exhibit 13, the impact of attending a rated site on the average number of monthly unexcused absences is shown with a diamond. These estimates are relatively more precisely estimated, and results in Figure A indicate that attending a rated ECEAP site is associated with a significant 15% decline in monthly absence rate. No such significant relationship exists between rated subsidy site attendance and absences (Figure B).

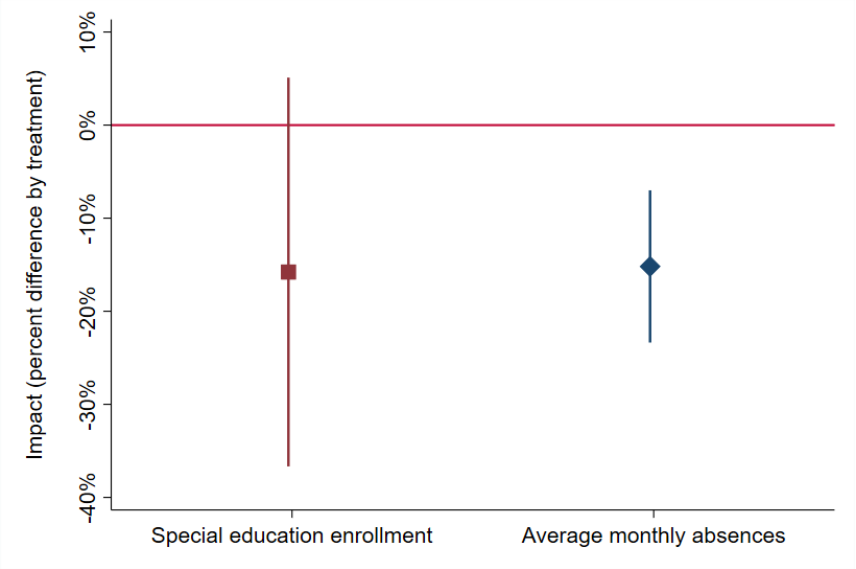
Summary. Overall, attending a rated pre-k site appears, in some contexts, to be modestly related to the outcomes examined here. However, these effects are apparent for children attending a rated ECEAP site, while they are null for children attending a rated subsidy site.[34]

[34] For subsidy sites we additionally investigate coaching as a potential moderator of the effects of pre-k rated status on child outcomes. Specifically, we were interested in whether a stronger effect would be present for sites receiving coaching, and sites receiving relatively more coaching. The results of this analysis were null, indicating that as we defined it, effects of being rated on child outcomes did not vary by coaching. Additional detail is presented in Appendix VI, Exhibit A21.

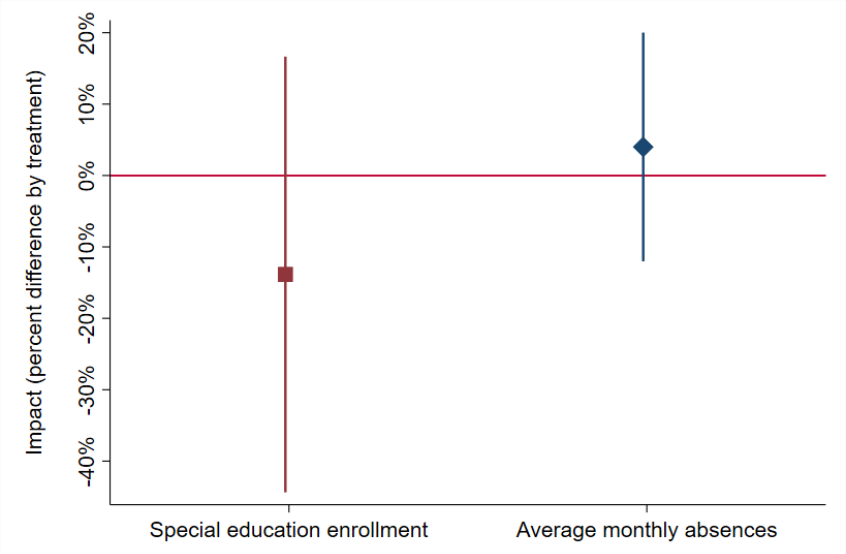
Exhibit 13

Pre-K Early Achievers Rated Status and Additional Kindergarten Outcomes

(A) ECEAP site rated status and kindergarten outcomes



(B) Subsidy site rated status and kindergarten outcomes



Impact: Percentage difference in outcome between the treated and comparison group.
Confidence interval: Range of values that likely includes the *true* treatment effect.

Notes:
Each point represents the impact of the marginal effect derived from a sperate regression model, and the lines represent the corresponding 90% confidence interval.
Each model controls for the full set of control variables and adjusts standard errors for clustering at the pre-k site level. Where WaKIDS outcomes are used the model additionally controls for the first year of implementation in school.

Question Two: Do Higher QRIS Ratings Predict Better Child Outcomes?

Method

To assess the relationship between rating level and child outcomes, we similarly apply an entropy weighting method now restricting our sample to children attending sites that have been rated.[35]

For this analysis, we primarily define treatment as "attending a site rated at-quality."[36] Most sites do not change their rating level within our sample period; therefore, we cannot implement a site-fixed effects approach.[37]

Kindergarten Readiness, Fall of Kindergarten. In Figure A of Exhibit 14, we explore the relationship between attending an ECEAP site rated at-quality and kindergarten readiness. Estimates suggest that children who attend an ECEAP site rated at-quality are 11% more likely to meet/exceed at least five domains on the WaKIDS assessment compared children who attend a site not rated at-quality, this estimate is significant at the 5% level. There is no significant difference in meeting all 6 domains between children who attend a site rated at-quality and those who do not.

In Figure B of Exhibit 14, we summarize results from parallel analysis of subsidy sites. Additionally, for subsidy sites our sample is large enough to examine the relationship between 1) rating at-quality and WaKIDS performance (denoted by the solid line in Figure B) and 2) rating greater than three versus rating at level three (denoted by the dashed line).

We observe that children who attend sites rated at-quality are significantly more likely to meet/exceed at least five domains (by 6%) and all six domains (by 10%) compared to children who attend sites not rated at-quality. Impact estimates are similar when comparing children who attend a site rated above level 3 to those who attend a site rated at level 3, although these estimates are imprecisely measured for meet/exceed all six domains.[38]

[35] Entropy weights are re-estimated depending on sample composition and how treatment is defined. For example, the entropy weights used when treatment was defined as "attending a rated site" will differ from those calculated when treatment is defined as "attending a site rated at-quality."

[36] For subsidy sites, we can additionally explore the predictive power of "attending a site rated greater than level three" (conditional on site being rated at-quality).

[37] For this analysis we estimate entropy weighted logistic and Poisson models that control for the full set of aforementioned covariates. In addition, we add the following time-invariant site-level controls: years-in-operation, CCA region, an indicator for initial rating, has been licensed (ECEAP), primary language ever non-English (ECEAP), received coaching (subsidy). All models are estimated to allow for clustering at the pre-k site-level.

[38] We re-estimated the relationship between subsidy site rating status and rating level with the inclusion of the 2018-2019 pre-k academic year—these results are summarized in Appendix VI, Exhibit A25. Results from this analysis are qualitatively similar to the primary results presented here although not statistically significant.

Kindergarten Readiness, Spring of Pre-K. Parallel analysis using this outcome, for ECEAP children only, indicates that attending an ECEAP site rated at-quality predicts a significant 9% increase in the probability of meeting/exceeding benchmark scores in at least five domains but no significant impact on the probability of meeting/exceeding on all six domains. These results are similar to the relationship between attending an ECEAP site rated at-quality and WaKIDS performance on at least five domains. Estimated results from this analysis can be found in Appendix V.

Additional Kindergarten Outcomes. Similar to our analysis regarding the impact of attending a rated site on special education enrollment and absence rates, attending an ECEAP site rated at-quality is associated with a sizable decrease in 1) the probability of special education enrollment in kindergarten and 2) the average number of monthly unexcused absences.

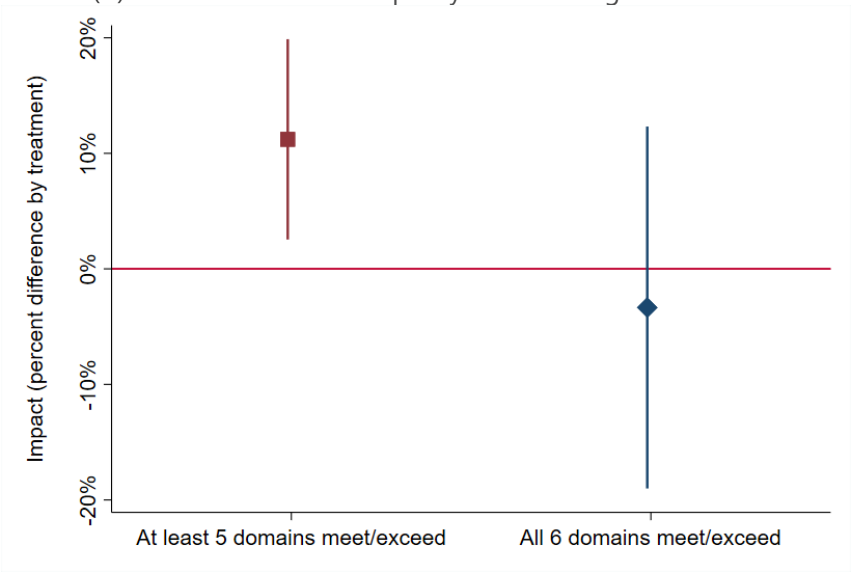
Results regarding the relationship between attending a subsidy site rated at (or above) quality and these outcomes are mixed and largely non-significant. Detailed results of these analyses are shown in Appendix VI.

Summary. Results largely support that attending a site rated at or above quality predicts greater kindergarten readiness for children in ECEAP and in child care centers receiving subsidies.

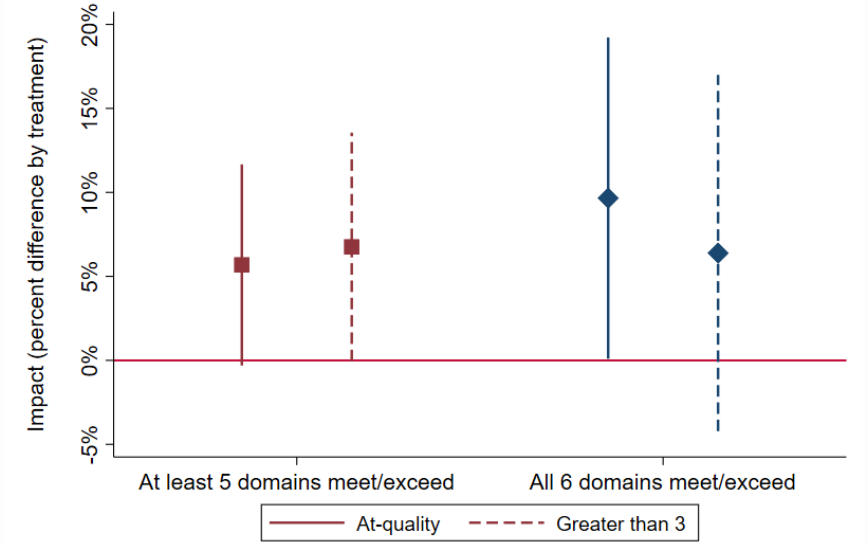
Exhibit 14

Pre-K Early Achievers Rating Level and Kindergarten Readiness

(A) ECEAP site rated "at-quality" and kindergarten outcomes



(B) Subsidy site rated at-quality or "above quality" and kindergarten outcomes



Impact: Percentage difference in outcome between the treated and comparison group.
Confidence interval: Range of values that likely includes the *true* treatment effect.

Notes:
Each point represents the impact of the marginal effect derived from a sperate regression model, and the lines represent the corresponding 90% confidence interval.
Each model controls for the full set of control variables and adjusts standard errors for clustering at the site level. Where WaKIDS outcomes are used the model additionally controls for the first year of implementation in school.

VII. Summary and Limitations

Results Summary

This report broadly explores the relationship between Early Achievers participation and rating level and outcomes for children enrolled in pre-k in the year prior to entering kindergarten. We find modest but not-significant associations with attending a rated ECEAP site and improvements in kindergarten readiness. Our findings suggest that rating completion largely has no relationship with improvements in kindergarten readiness and other kindergarten outcomes for subsidy sites.

With respect to rating levels, we find that enrollment in pre-kindergarten sites rating at (or above) quality is a modest positive predictor of greater kindergarten readiness for children in ECEAP sites and those with child care subsidy. Our analysis revealed no significant differences in associations across racial/ethnic groups.

Limitations

The major limitation of this study is the inability to randomly assign sites and children to QRIS program participation. A random assignment would increase our confidence that the group differences we estimate are due to rating completion and corresponding rating level and not due to other unobserved characteristics of children or to alternate policies that impact similar outcomes and went into effect around the same time as the Early Start Act.

At the child-level, we are concerned that children with unobserved characteristics—such as greater resources or higher cumulative risk—are most likely to select into sites that have completed an Early Achievers rating and rate at higher levels. That means there is a possibility that selection into early learning sites could be driving outcomes rather than program participation itself. At the site level, it is possible that other policies or circumstances were changing at the same time rating occurs, and it is these policies that are driving our observed changes in outcomes rather than Early Achievers (e.g., change in the composition of staff coincident with EA or the adoption of new curriculum).

Our rigorous empirical research strategy takes strides to alleviate these concerns, and our sensitivity analysis further supports our main results. However, we ultimately cannot fully rule out the possibility that decisions or circumstances surrounding site participation and children's enrollments drive the outcomes we observe, rather than quality improvement itself.

The interpretability and generalizability of our results are further limited by the fact that our comparison groups throughout the analysis are relatively small. For example, for ECEAP sites, 97% of the sample attends a site that rates at-quality and only 3% attend a site that rates below quality. For subsidy sites, 83% of the sample attends sites that rate above quality and only 17% attend a site that rates below quality. This lack of variation in treatment status within the sample leads to some uncertainty regarding our estimated findings and severely limits our ability to explore how the relationship between rating level and child outcomes varies across subgroups such as region, race/ethnicity, sex, and other potential variables of interest.

Another limitation presented by the data regards the TS Gold assessment of kindergarten readiness, used for ECEAP students in the spring of the pre-k year. In particular, the possible range of points and benchmark threshold values have undergone major changes, especially around the time most sites completed rating (See Exhibit A15 in Appendix IV). One concern is that our predicted outcomes reflect (or are masked by) administrative scoring changes and *not* changes in actual kindergarten readiness related to site-level quality improvement or rating level. Considering this uncertainty, we caution against drawing conclusive inferences based on analysis using the TS Gold.

A final point about drawing policy-relevant inference from this study, or others in the literature, regards the use of overall quality ratings as predictors. These ratings reflect site-level quality in broad and non-specific ways. Ratings are derived from points aggregated across standard areas, with some standard areas—in particular Learning Environment and Interactions—expected to be more directly tied to child outcomes than others.[39] Further, the Early Achievers system allows for sites to receive a quality rating without being evaluated on all standard areas, and programs can earn higher ratings (i.e., above what is required for an at-quality rating) through different means. From a research perspective, this may mask true underlying variation in quality and likely attenuates estimated associations with child outcomes.

More discussion regarding our methodological approach and the limitations presented by the research design and data can be found in Appendix V and VII.

Future Work

WSIPP is directed to submit additional Early Achievers reports in December 2021 and 2022. Our 2022 report will address 3rd-grade achievement test outcomes for cohorts in the present study,[40] as well as a benefit-cost analysis.

Several research questions of interest for our December 2021 report are discussed below. Additionally, within the constraints of available data, we have some flexibility to address questions emerging from the present report or other sources.

Quality Standard Areas

As noted throughout this report, the limited variation in overall quality ratings constrains our analyses and conclusions. In our 2021 report, to the extent possible given the data, we will investigate variation within quality standard areas to predict child outcomes. In particular, variation in scores on "Learning Environment and Interactions"—which is the common component of all EA quality ratings—may prove equally or more informative than overall quality ratings. This reflects our third guiding research question.

[39] Fox et al. (2019).

[40] School closures, remote learning, and adjustments to testing protocols due to COVID-19 may require adjustments.

Quality of Care for Younger Children

In the present report, we focused on quality in four-year-old pre-kindergarten children.[41] However, Early Achievers is intended to improve child care quality for children of all ages. If the data allows, we will examine infants' and toddlers' child care quality experiences and outcomes.[42]

Finally, we note that WSIPP's plans for the evaluation of Early Achievers in 2021 and 2022 may require adjustment to accommodate changes to programs and available data due to closures related to COVID-19. Early Achievers on-site evaluations were put on hold in March 2020; delaying completion of site ratings that would have been included in our AY 2019-20 ECEAP and child care cohorts. Additionally, we understand that there has been instability in child care availability, enrollment, and attendance, and a required shifting of resources, as a result of the COVID-19 pandemic. The March 2020 shift to remote learning for public elementary school students impacts available data. Broader social and economic impacts must also be considered in terms of their implications for children's outcomes.

[41] This approach is consistent with the existing evaluation of QRIS and child outcomes. In our 2019 review, very few studies had reported on QRIS and child outcomes for children prior to pre-kindergarten.

[42] We anticipate being able to observe three cohorts of children in early care during the Early Achievers roll-out, corresponding with kindergarten enrollment and outcomes data in AY 2017-18 through 2019-20.

Acknowledgments

The authors would like to thank staff at DCYF, ERDC, Child Care Aware of Washington, and Cultivate Learning at the University of Washington, for their assistance in understanding Early Achievers and the child care and early learning landscape in Washington State and for their support in obtaining administrative data to facilitate WSIPP's evaluation. We also thank staff at OSPI and DOH for contributing data to WSIPP's evaluation and their guidance in data use. In particular, Warren Wessling, Kevin Cummings, Joyce Kilmer, Sara Schwartz-Jewell, and Rachael Brown-Kendall at DCYF; Tom Aldrich, Tim Norris, and Jeffrey Thayne at ERDC; Karen Sampson and Sarah Kelley at Child Care Aware; Johnna Lee, DeEtta Simmons, and Gail Joseph at Cultivate Learning; and Lucas Snider at OSPI have graciously answered many questions about programs, data, and WSIPP's evaluation. Vickie Ybarra, Kevin Cummings, and Dani Fumia provided helpful commentary on earlier drafts of this report. Finally, we thank Chasya Hoagland, John Hansen, and Anthony Ellis for their contributions to this work.

Appendix

Early Achievers Evaluation Report Two: *Pre-Kindergarten Quality and Child Outcomes in Kindergarten*

Appendices

I. Data Sources

Exhibit A1 details data sources utilized for the Early Achievers evaluation. Data coverage indicates dates for all data requested. In the present report, our samples reflect a narrower set of cohorts selected to best address the association of pre-kindergarten quality and children's kindergarten outcomes.

Exhibit A1

Data Sources used for Early Achievers Evaluation

Data type	Data systems or reports	Data source	Coverage
Site-level data			
Early Achievers ratings and rating dates	WELS	DCYF	All sites receiving an initial rating, re-rating, or renewal rating in Early Achievers from Jul. 2012 – Apr. 2020.
Early Achievers participation milestone dates	Early Achievers Private Pay Monitoring Report; MERIT Reports	DCYF CCA of WA	All participation milestone dates in monthly MERIT reports (Aug. 2012 - Mar. 2016) and monthly Early Achievers Private Pay Monitoring Reports (Apr. 2016 – Apr. 2020).
Early Achievers consultation and coaching dates	CCA system	CCA of WA DCYF	Pre-rating consultation dates for all licensed facilities working with CCA from Jul. 2012 through Apr. 2020.
ECEAP site characteristics	ELMS	DCYF/ERDC	All ECEAP sites serving children from AY 2013-14 – AY 2019-20.
Licensed child care facility characteristics	FamLink; WA Compass	DCYF CCA of WA	All licensed child care sites with an active license from Sep. 2009 – Aug. 2020.
Child-level data			
Child ECEAP eligibility and enrollments; TS-Gold assessments and dates	ELMS	DCYF/ERDC	All children enrolled in ECEAP from AY 2013-14 – AY 2019-20.
Child care subsidy participation	SSPS	DCYF	All children receiving child care subsidy through WCCC, SCC, or child welfare from Sep. 2009 – Aug. 2020.
Child health at birth; time-varying family characteristics	Birth statistical files	DOH	All live births from Sep. 2008 – Aug. 2016.
Child K-3 program participation and assessment data	CEDARS	OSPI/ERDC	All K-3 children from AY 2014-15 – AY 2020-21 who match an individual identified for ECEAP or child care subsidy in target years.

Notes:
AY = Academic year.
WELS = Web-based Early Learning System.
DCYF = Department of Children, Youth, and Families.
MERIT= Managed Education and Registry Information Tool.
ELMS = Early Learning Management System.
ERDC = Education Research & Data Center.
SSPS = Social Service Payment System.
CCA = Child Care Aware.
DOH = Department of Health.
CEDARS = Comprehensive Education Data and Research System.
OSPI = Office of the Superintendent of Public Instruction.

II. Sample Construction and Attrition

In this section, we provide additional detail regarding study sample construction for sites and children in ECEAP and subsidized child care. This process consisted of two major steps: foundational sample construction and analysis sample construction. These steps were completed separately for ECEAP and child care subsidy sites and enrolled children. The foundational samples will be utilized in the remaining WSIPP Early Achievers evaluation reports. The analysis samples apply site and child inclusion and exclusion criteria specific to the research questions addressed in the present report.

Foundational Sample

We first integrated site-level Early Achievers records from multiple source systems and reports over time to construct a site by academic year (AY) crosswalk file. This crosswalk currently includes sites that were present in Early Achievers administrative records from the 2012-13 AY through the 2018-19 AY. When data are mature, we will add one additional cohort of Early Achievers records, covering activity in the 2019-20 AY. Site-level data sources for Early Achievers participation and ratings include monthly MERIT demographic, registration, and evaluation request reports; Early Achievers Monitoring Reports; and web-based Early Learning System (WELS) Early Achievers ratings records. Additionally, to observe site characteristics not included in Early Achievers records; we incorporated licensing data and ECEAP site data. For non-ECEAP licensed child care, sources included FamLink monthly reports and WA Compass records. ECEAP site characteristics were pulled from the Early Learning Management System (ELMS) site data. This crosswalk provides site-level information about Early Achievers participation and ratings, as well as other characteristics, for each year in our study period.

Second, we used child enrollment records to identify all children who participated in ECEAP or subsidized child care during our target years. ECEAP enrollment records were used to define ECEAP participation. Monthly payment records from the Social Service Payment System (SSPS) were used to define participation in subsidized child care for each AY.

Third, we linked child enrollment records to each site attended, for each year of attendance. Children can attend multiple child care sites, and each site is observed here. This resulted in a child by site by AY panel in which we identify each year and site of a child's child care and early learning (CC/EL) attendance, and critically for the purpose of this evaluation, observe Early Achievers treatment status for each year and site of a child's attendance. Imperfect historic record systems result in the loss of both sites and children in this step. Specifically, not all sites in the crosswalk have observed child enrollments, and not all children in enrollment records could be accurately matched to a site given the available site identifiers. Children in subsidized child care during our target years who were born before September 2008 or after August 2015 were not part of our defined study cohorts and were excluded in this step.

Analysis Samples

All additional steps in sample construction were taken to specify an analysis sample for the specific research questions addressed in the present report.

Starting with the foundational child by site by AY panel, we restricted our sample to children observed in ECEAP or subsidy and then in kindergarten enrollment data during the 2014-15 through 2018-19 AYs. Children may have multiple enrollments in their pre-k year, both simultaneous and consecutive. We selected children's first enrollment site during the AY prior to their first kindergarten enrollment and refer

to this as the child's *pre-k enrollment*.[43] To accurately assign children to a single "treatment" group (i.e., rated or unrated pre-k enrollment site, quality rating level of the pre-k enrollment site) based on their pre-k enrollment, we excluded children whose first pre-k enrollment occurred in more than one site simultaneously. This exclusion applied only to the subsidy sample, as children may not simultaneously enroll in multiple ECEAP sites.

We then applied additional restrictions to the site and child study samples. We omitted children who were enrolled in pre-k during the 2013-14 AY, and kindergarten during the 2014-15 AY.[44] Additionally, due to concerns about missing covariate data, our primary analyses omitted children enrolled in pre-k during the 2018-19 AY, and in kindergarten during the 2019-20 AY, although for subsidy participants we test the sensitivity of results to the inclusion of this cohort (see Appendix VI Exhibit A25).

We omitted sites with the following characteristics:

- Missing or illogical Early Achievers milestone date records;
- No observed subsidy enrollments in SSPS records after 2015 or before 2018;
- Enrollments observed for less than two years in SSPS records;
- Sites with fewer than four children enrolled;
- Licensed sites that started an ECEAP contract in the same period as Early Achievers engagement. We consider starting an ECEAP contract to be a confounding "treatment" that would preclude attributing any potential effects to Early Achievers. Further, because eligibility criteria for ECEAP differ from those for child care subsidy, we were also concerned about changes in the composition of children attending these sites before and after an ECEAP contract, which would invalidate analysis of within-site change in child outcomes pre and post Early Achievers rating.[45]

We excluded children with the following characteristics:

- Subsidy and ECEAP enrollment simultaneously in pre-k and
- Exposure to a rated site at any time *after* pre-k enrollment (including after starting kindergarten) but not *during* pre-k enrollment.

The second child exclusion criteria above was applied because if children were untreated (in an unrated site) in their pre-k enrollment, but were actually exposed to treatment during the period in which outcomes are assessed, their inclusion in the control group would attenuate potential treatment effects.

Our ECEAP analysis sample comprised 16,384 children in 297 sites. Our final pre-k subsidy analysis sample comprised 14,008 children in 1,106 sites. When we retain only sites with an observed Early Achievers rating, our ECEAP sample includes 12,290 children in 288 sites, and our subsidy sample includes 6,071 children in 519 sites.

[43] In this step we selected children observed in kindergarten enrollment records in our target years and in the year prior to their first kindergarten enrollment, without regard to year of expected kindergarten enrollment based on birth cohort. Of children observed in a pre-k setting at age four, about 90% of the ECEAP sample, and about 72% of the subsidy sample, was also observed in kindergarten. Kindergarten enrollment before or after the "expected" year was rare in our sample. For ECEAP 0.2% of children appeared to be enrolled in kindergarten one year earlier than expected based on birth cohort. For subsidy 0.5% appeared to be enrolled one year earlier than expected, and 2% appeared to be enrolled one year later than expected based on birth cohort.
[44] This restriction is motivated by the fact that WaKIDS and TS GOLD data from the 2014-15 kindergarten year are not comparable to those reported in other years. We carry this restriction throughout all outcome analyses for sample consistency in order to facilitate comparable inferences across outcomes.
[45] DCYF. Getting help paying for child care. *Child care subsidy programs.*

Exhibit A2

Comparing Full Pre-K to Kindergarten Sample to Analysis Sample—
ECEAP Child Characteristics

Characteristic	Full pre-k to K sample	Analysis sample
Funding model:		
Part-day	0.87	0.86
	(0.34)	(0.34)
School-day	0.11	0.12
	(0.31)	(0.32)
Work-day	0.03	0.03
	(0.17)	(0.16)
Female	0.50	0.51
	(0.50)	(0.50)
Race:		
Black	0.09	0.07
	(0.28)	(0.26)
Hispanic/Latinx	0.43	0.45
	(0.50)	(0.50)
Other	0.14	0.13
	(0.35)	(0.33)
White	0.34	0.35
	(0.47)	(0.48)
Primary language:		
English	0.62	0.62
	(0.49)	(0.48)
Spanish	0.29	0.31
	(0.46)	(0.46)
Other	0.09	0.07
	(0.28)	(0.26)
In ECEAP in previous year	0.26	0.37
	(0.44)	(0.48)
Moved to current site	0.06	0.06
	(0.23)	(0.23)
Simultaneously in ECEAP site and kindergarten	0.13	0.04
	(0.34)	(0.19)
Full-time kindergarten	0.91	0.97
	(0.29)	(0.18)
At least 5 domains meet/exceed Fall TS Gold	0.12	0.21
	(0.32)	(0.41)

Exhibit A2, Continued

Comparing Full Pre-K to Kindergarten Sample to Analysis Sample—
ECEAP Child Characteristics

Characteristic	Full pre-k to K sample	Analysis sample
Mother married	0.40	0.44
	(0.49)	(0.50)
Mother's education at birth:		
Less than high school	0.36	0.34
	(0.48)	(0.47)
High school complete	0.34	0.35
	(0.47)	(0.48)
More than high school	0.30	0.31
	(0.46)	(0.46)
Mother's age at birth	26.02	26.47
	(5.95)	(6.00)
First born child	0.28	0.26
	(0.45)	(0.44)
Premature birth	0.08	0.08
	(0.27)	(0.27)
Number of observations	31,435	10,278

Exhibit A3

Comparing Full Pre-K to Kindergarten Sample to Analysis Sample—Kindergarten School Characteristics for ECEAP Students

Characteristic	Full pre-k to K sample	Analysis sample
Percent female	0.48	0.48
	(0.03)	(0.03)
Percent American Indian/Alaskan Native	0.01	0.01
	(0.04)	(0.04)
Percent Asian	0.05	0.05
	(0.07)	(0.07)
Percent Black	0.06	0.05
	(0.09)	(0.09)
Percent Hispanic/Latinx	0.34	0.34
	(0.25)	(0.25)
Percent white	0.45	0.46
	(0.25)	(0.25)
Percent low-income	0.66	0.66
	(0.21)	(0.20)
Percent section 504	0.02	0.02
	(0.02)	(0.02)
Percent diagnosed disabled	0.16	0.16
	(0.05)	(0.05)
Percent students enrolled K-3rd grade	0.65	0.65
	(0.12)	(0.13)
Percent foster care	0.01	0.01
	(0.01)	(0.01)
Percent unhoused	0.04	0.04
	(0.04)	(0.04)
Number of observations	31,155	10,278

Exhibit A4

Comparing Full Pre-K to Kindergarten Sample to Analysis Sample—Subsidy Child Characteristics

Characteristic	Full pre-k to K sample	Analysis sample
Female	0.50	0.51
	(0.50)	(0.50)
Race:		
Black	0.12	0.09
	(0.32)	(0.28)
Hispanic	0.32	0.31
	(0.47)	(0.46)
Other	0.16	0.17
	(0.36)	(0.38)
White	0.40	0.43
	(0.49)	(0.50)
Primary language:		
English	0.79	0.87
	(0.40)	(0.34)
Spanish	0.15	0.11
	(0.36)	(0.32)
Other	0.06	0.02
	(0.23)	(0.15)
Number of subsidy sites previously attended:		
0	0.41	0.40
	(0.49)	(0.49)
1	0.27	0.28
	(0.44)	(0.45)
2+	0.32	0.33
	(0.47)	(0.47)
In subsidy care for at least one year prior	0.40	0.47
	(0.49)	(0.50)
Moved to current site	0.52	0.55
	(0.49)	(0.49)
Simultaneously in subsidy site and kindergarten	0.53	0.47
	(0.50)	(0.50)
Full-time kindergarten	0.90	0.97
	(0.30)	(0.16)

Exhibit A4, Continued

Comparing Full Pre-K to Kindergarten Sample to Analysis Sample—
Subsidy Child Characteristics

	Full pre-k to K sample	Analysis sample
Mother married	0.27	0.26
	(0.45)	(0.44)
Mother's education at birth:		
Less than high school	0.29	0.26
	(0.46)	(0.44)
High school complete	0.36	0.37
	(0.48)	(0.48)
More than high school	0.35	0.37
	(0.48)	(0.48)
Mother's age	24.75	24.72
	(5.49)	(5.37)
First born child	0.32	0.33
	(0.47)	(0.47)
Premature birth	0.08	0.08
	(0.27)	(0.27)
Number of observations	29,802	9,185

Exhibit A5

Comparing Full Pre-K to Kindergarten Sample to Analysis Sample—Kindergarten School Characteristics for Subsidy Sample

	Full pre-k to K sample	Analysis sample
Percent female	0.48	0.48
	(0.03)	(0.02)
Percent American Indian/Alaskan Native	0.01	0.01
	(0.03)	(0.03)
Percent Asian	0.06	0.05
	(0.08)	(0.07)
Percent Black	0.06	0.06
	(0.10)	(0.08)
Percent Hispanic/Latinx	0.31	0.30
	(0.24)	(0.23)
Percent white	0.45	0.47
	(0.24)	(0.24)
Percent low-income	0.63	0.63
	(0.22)	(0.20)
Percent section 504	0.02	0.03
	(0.02)	(0.02)
Percent diagnosed disabled	0.16	0.16
	(0.05)	(0.05)
Percent students enrolled K-3rd grade	0.66	0.65
	(0.11)	(0.11)
Percent foster care	0.01	0.01
	(0.01)	(0.01)
Percent unhoused	0.04	0.04
	(0.03)	(0.03)
Number of observations	29,611	9,153

III. Descriptive Statistics

Exhibit A6

ECEAP Site-Level Characteristics (2015-2018)

Characteristic	All sites (1)	Analysis sample (2)	Analysis sample, rated (3)
Number of sites	469	282	269
First-year of observed child enrollment	2014	2014	2014
Last year of observed child enrollment	2018	2018	2018
Years in operation	3.88	4.4	4.47
	(1.43)	(0.94)	(0.86)
Child care center	0.98	1.00	1.00
	(0.12)	(0)	(0)
CCA Region:			
Central	0.13	0.15	0.15
	(0.34)	(0.35)	(0.35)
Eastern	0.22	0.21	0.2
	(0.41)	(0.41)	(0.40)
King and Pierce	0.26	0.26	0.29
	(0.44)	(0.44)	(0.46)
Northwest	0.13	0.12	0.12
	(0.33)	(0.33)	(0.32)
Olympic Peninsula	0.11	0.13	0.11
	(0.31)	(0.33)	(0.31)
Southwest	0.15	0.13	0.13
	(0.36)	(0.34)	(0.34)
Director primary language ever non-English	0.07	0.08	0.08
	(0.25)	(0.26)	(0.27)
Director secondary language ever non-English	0.48	0.49	0.5
	(0.49)	(0.49)	(0.49)
Is licensed	0.34	0.26	0.3
	(0.48)	(0.44)	(0.46)
ECEAP enrollment slots	37.94	41.49	42.4
	(34.35)	(35.36)	(31.84)
Rated	0.81	0.85	1.00
	(0.39)	(0.36)	(0)
Initial rating at-quality	0.72	0.79	0.8
	(0.45)	(0.41)	(0.4)
Number of observations (site x year)	1,386	963	721

Notes:
Column (2) omits children who are not in the WA Kids database, sites that are in operation for less than one year, and observations missing in the set of control variables (controls come from CEDARS, OSPI, DOH, & Census).
Column (3) additionally restricts the sample to rated site-year observations (i.e., sites that are treated in the post-treatment period).
Column (2) corresponds to our baseline analytical sample, and Column (3) corresponds to the sample used to evaluate questions 2 and 3.

Exhibit A7

ECEAP Census Tract Community Characteristics (2015-2018)

Characteristic	All sites (1)	Analysis sample (2)	Analysis sample, rated (3)
% Pop 25 years and up – Less than 9th grade	6.3	6.76	6.89
	(8.36)	(8.83)	(9.07)
% Pop 25 years and up – High school graduate or higher	86.5	85.74	85.56
	(11.22)	(11.67)	(11.97)
% Pop 25 years and up – Bachelor's degree or higher	23.91	22.55	22.91
	(13.57)	(12.55)	(12.71)
Unemployment rate	7.71	7.76	7.31
	(3.80)	(3.66)	(3.50)
Log median household income	10.89	10.88	10.9
	(0.35)	(0.33)	(0.34)
Percent families below FPL	11.66	11.88	11.92
	(8.46)	(8.49)	(8.73)
Percent American Indian and Alaska Native	4.27	4.34	4.44
	(7.30)	(7.46)	(7.86)
Percent Asian	7.31	7.13	7.61
	(9.53)	(9.43)	(9.89)
Percent Black or African American	5.41	5.29	5.62
	(7.11)	(7.15)	(7.31)
Percent Hispanic/Latinx	16.31	17.25	17.55
	(18.67)	(19.27)	(19.63)
Percent Native Hawaiian and other Pacific Islander	1.32	1.27	1.34
	(2.18)	(2.19)	(2.23)
Percent other race	6.33	6.75	6.9
	(9.41)	(10.14)	(10.29)
Percent white	81.11	80.94	79.97
	(16.7)	(17.23)	(17.78)
Number of observations (site x year)	1,386	963	721

Notes:
Column (2) omits children who are not in the WA Kids database, sites that are in operation for less than one year, and observations missing in the set of control variables (controls come from CEDARS, OSPI, DOH, & Census).
Column (3) additionally restricts the sample to rated site-year observations (i.e., sites that are treated in the post-treatment period).
Column (2) corresponds to our baseline analytical sample, and Column (3) corresponds to the sample used to evaluate questions 2 and 3.

Exhibit A8

Subsidy Site-Level Characteristics (2015-2018)

Characteristic	All sites (1)	Analysis sample (2)	Analysis sample, rated (3)
Number of sites	4,362	732	326
First year of observed child enrollment	2011.9	2010.83	2010.76
	(2.50)	(1.53)	(1.41)
Last year of observed child enrollment	2019.34	2019.78	2019.71
	(1.16)	(0.59)	(0.68)
Years in operation	7.45	8.95	8.95
	(2.80)	(1.64)	(1.56)
Child care center	0.34	0.78	1.00
	(0.47)	(0.42)	(0)
License capacity	31.19	65.87	85.53
	(39.85)	(46.87)	(41.83)
Proportion subsidy enrolled	0.47	0.57	0.48
	(0.30)	(0.28)	(0.27)
CCA Region:			
Central	0.19	0.19	0.07
	(0.39)	(0.39)	(0.25)
Eastern	0.13	0.14	0.18
	(0.34)	(0.35)	(0.38)
King and Pierce	0.40	0.33	0.4
	(0.49)	(0.47)	(0.49)
Northwest	0.13	0.13	0.11
	(0.33)	(0.34)	(0.31)
Olympic Peninsula	0.09	0.12	0.12
	(0.29)	(0.33)	(0.32)
Southwest	0.06	0.09	0.13
	(0.24)	(0.28)	(0.34)
Director primary language ever non-English	0.28	0.15	0.03
	(0.45)	(0.36)	(0.17)
Director pecondary language ever non-English	0.34	0.43	0.49
	(0.47)	(0.50)	(0.50)
Received coaching	0.64	0.71	0.81
	(0.48)	(0.45)	(0.39)
Initial rating at-quality	0.84	0.81	0.76
	(0.37)	(0.39)	(0.43)
Rated	0.38	0.50	1.00
	(0.49)	(0.50)	(0)
Number of observations (site x year)	16,873	2,359	943

Notes:
Column (2) omits children who are not in the WA Kids database, sites that are in operation for less than one year, and observations missing in the set of control variables (controls come from CEDARS, OSPI, DOH, & Census).
Column (3) additionally restricts the sample to rated site-year observations (i.e., sites that are treated in the post-treatment period).
Column (2) corresponds to our baseline analytical sample, and Column (3) corresponds to the sample used to evaluate questions 2 and 3.

Exhibit A9

Subsidy Census Tract Community Characteristics (2015-2018)

Characteristic	All sites (1)	Analysis sample (2)	Analysis sample, rated (3)
% Pop 25 years and up – Less than 9th grade	8.41	7.56	5.02
	(10.13)	(9.31)	(5.75)
% Pop 25 years and up – High school graduate or higher	84.25	84.86	88.37
	(13.29)	(12.52)	(8.65)
% Pop 25 years and up – Bachelor's degree or higher	25.99	23.9	27.11
	(15.34)	(13.61)	(14.71)
Percent estimate unemployment rate	7.29	7.64	7.07
	(3.49)	(3.78)	(3.53)
Log median household income	10.95	10.86	10.89
	(0.35)	(0.35)	(0.37)
Percent estimate families below FPL	11.78	12.64	11.75
	(8.65)	(9.07)	(9.04)
Percent American Indian and Alaska Native	3.28	3.31	3.30
	(5.12)	(3.34)	(3.41)
Percent Asian	10.64	8.82	10.04
	(11.55)	(9.66)	(10.30)
Percent Black or African American	7.36	6.71	7.21
	(9.17)	(7.70)	(7.48)
Percent Hispanic/Latinx	20.92	20.38	14.14
	(22.88)	(21.85)	(14.50)
Percent Native Hawaiian and other Pacific Islander	1.28	1.28	1.38
	(2.02)	(1.93)	(1.96)
Percent other race	8.15	7.54	5.42
	(10.36)	(9.53)	(6.70)
Percent white	75.45	78.8	79.43
	(18.23)	(15.51)	(15.45)
Number of observations (site x year)	16,873	2,359	943

Notes:
Column (2) omits children who are not in the WA Kids database, sites that are in operation for less than one year, and observations missing in the set of control variables (controls come from CEDARS, OSPI, DOH, & Census).
Column (3) additionally restricts the sample to rated site-year observations (i.e., sites that are treated in the post-treatment period).
Column (2) corresponds to our baseline analytical sample, and Column (3) corresponds to the sample used to evaluate questions 2 and 3.

Exhibit A10

Child-level characteristics, ECEAP-site attendees (2015-2018)

Characteristic	Full sample	Attending rated sites
Attends site post-rating complete	0.76	
	(0.42)	
Funding model:		
Part-time	0.86	0.85
	(0.34)	(0.36)
School day	0.12	0.13
	(0.32)	(0.34)
Work day	0.03	0.03
	(0.16)	(0.16)
Female	0.51	0.51
	(0.50)	(0.50)
Race/ethnicity:		
Black	0.07	0.08
	(0.26)	(0.27)
Hispanic/Latinx	0.45	0.46
	(0.50)	(0.50)
Other	0.13	0.13
	(0.33)	(0.34)
White	0.35	0.33
	(0.48)	(0.47)
Primary language:		
English	0.62	0.61
	(0.48)	(0.49)
Spanish	0.31	0.31
	(0.46)	(0.46)
Other	0.07	0.08
	(0.26)	(0.26)
In ECEAP in previous year	0.37	0.41
	(0.48)	(0.49)
Moved to current site	0.06	0.05
	(0.23)	(0.21)
Simultaneously in subsidy site and kindergarten	0.04	0.03
	(0.19)	(0.18)
Full-time kindergarten	0.97	0.97
	(0.18)	(0.17)
At least 5 domains meet/exceed, Fall TS Gold	0.21	0.22
	(0.41)	(0.41)

Note:
Table reports unweighted sample averages.

Exhibit A10, Continued

Child-level Characteristics, ECEAP-site Attendees (2015-2018)

Characteristic	Full sample	Attending rated sites
Mother married	0.44	0.44
	(0.50)	(0.50)
Mother's education at birth:		
Less than high school	0.34	0.33
	(0.47)	(0.47)
High school complete	0.35	0.35
	(0.48)	(0.48)
More than high school	0.31	0.32
	(0.46)	(0.47)
Mother's age at birth	26.47	26.60
	(6.00)	(6.03)
First born child	0.26	0.26
	(0.44)	(0.44)
Premature birth	0.08	0.07
	(0.27)	(0.26)
Number of observations (child x year)	10,278	7,868

Note:
Table reports unweighted sample averages.

Exhibit A11

Kindergarten School-level Characteristics, ECEAP-site Attendees (2015-2018)

Characteristic	Full sample	Attending rated sites
Percent female	0.48	0.48
	(0.03)	(0.03)
Percent American Indian/ Alaskan Native	0.01	0.01
	(0.04)	(0.04)
Percent Asian	0.05	0.05
	(0.07)	(0.07)
Percent Black	0.05	0.06
	(0.09)	(0.09)
Percent Hispanic/Latinx	0.34	0.34
	(0.25)	(0.25)
Percent white	0.46	0.45
	(0.25)	(0.25)
Percent low-income	0.66	0.66
	(0.20)	(0.20)
Percent section 504	0.02	0.02
	(0.02)	(0.02)
Percent diagnosed disabled	0.16	0.16
	(0.05)	(0.05)
Percent students enrolled K-3rd grade	0.65	0.65
	(0.13)	(0.13)
Percent foster care	0.01	0.01
	(0.01)	(0.01)
Percent unhoused	0.04	0.04
	(0.04)	(0.04)
Number observations (child x year)	10,278	7,844

Note:
Table reports unweighted sample averages.

Exhibit A12

Child-level Characteristics, Subsidy-site Attendees (2015-2018)

Characteristic	Full sample	Attending rated sites
Attends site post-rating complete	0.52	
	(0.50)	
Female	0.51	0.50
	(0.50)	(0.50)
Race/ethnicity:		
Black	0.09	0.09
	(0.28)	(0.29)
Hispanic/Latinx	0.31	0.27
	(0.46)	(0.44)
Other	0.17	0.18
	(0.38)	(0.39)
White	0.43	0.45
	(0.50)	(0.50)
Primary language:		
English	0.87	0.90
	(0.34)	(0.30)
Spanish	0.11	0.07
	(0.32)	(0.26)
Other	0.02	0.02
	(0.15)	(0.15)
Number of subsidy sites previously attended:		
0 sites	0.40	0.39
	(0.49)	(0.49)
1 site	0.28	0.27
	(0.45)	(0.44)
2+ sites	0.33	0.34
	(0.47)	(0.47)
In subsidy care for at least one year prior	0.47	0.51
	(0.50)	(0.50)
Moved to current site	0.55	0.54
	(0.49)	(0.49)
Simultaneously in subsidy site and kindergarten	0.47	0.43
	(0.50)	(0.49)
Full-time kindergarten	0.97	0.98
	(0.16)	(0.14)

Exhibit A12, Continued

Child-level Characteristics, Subsidy-site Attendees (2015-2018)

Characteristics	Full sample	Attending rated sites
Mother married	0.26	0.25
	(0.44)	(0.43)
Mother's education at birth:		
Less than high school	0.26	0.23
	(0.44)	(0.42)
High school complete	0.37	0.37
	(0.48)	(0.48)
More than high school	0.37	0.40
	(0.48)	(0.49)
Mother's age at birth	24.72	24.76
	(5.37)	(5.34)
First born child	0.33	0.34
	(0.47)	(0.48)
Premature birth	0.08	0.08
	(0.27)	(0.28)
Number of observations (child x year)	9,153	4,380

Note:
Table reports unweighted sample averages.

Exhibit A13

Kindergarten School-level Characteristics, Subsidy-site Attendees (2015-2018)

Characteristic	Full sample	Attending rated sites
Percent female	0.48	0.48
	(0.02)	(0.03)
Percent American Indian/Alaskan Native	0.01	0.01
	(0.03)	(0.02)
Percent Asian	0.05	0.06
	(0.07)	(0.08)
Percent Black	0.06	0.06
	(0.08)	(0.08)
Percent Hispanic/Latinx	0.30	0.26
	(0.23)	(0.19)
Percent white	0.47	0.49
	(0.24)	(0.23)
Percent low-income	0.63	0.61
	(0.20)	(0.21)
Percent section 504	0.03	0.03
	(0.02)	(0.02)
Percent diagnosed disabled	0.16	0.16
	(0.05)	(0.05)
Percent foster care	0.01	0.01
	(0.01)	(0.01)
Percent unhoused	0.04	0.04
	(0.03)	(0.03)
Percent students enrolled K-3rd grade	0.65	0.64
	(0.11)	(0.10)
Number of observations (child x year)	9,153	4,380

Note:
Table reports unweighted sample averages.

IV. Kindergarten Readiness Assessment

The Washington Kindergarten Inventory of Developing Skills (WaKIDS) is utilized in Washington State public elementary schools to assess children's kindergarten readiness. The WaKIDS is a custom version of the Teaching Strategies GOLD™ developmental assessment (TSG); WaKIDS items represent a subset of objectives and underlying dimensions from the full TSG. The WaKIDS observational assessment is completed by children's teachers in the first two months of children's kindergarten year. In addition, ECEAP sites assess children's development using the full TSG in the Fall and Spring of each academic year.

The TSG/WaKIDS is a snapshot record of teachers' observations of children's knowledge, skills, and abilities in six domains: social-emotional, physical, cognitive, language, literacy, and mathematics. For each objective and dimension on the assessment, teachers observe children's behavior over a period of time and assign a score indicating where children's demonstrated knowledge and behaviors fall on a developmental continuum. Item scores are then summed across items on a domain. Children are considered to meet "widely held expectations" in a given domain if they meet or exceed a benchmark score indicating that they have acquired a set of age-appropriate skills. Kindergarten readiness is considered for each domain as meeting or exceeding a benchmark score indicating consistent demonstration of the knowledge, skills, and abilities that would be expected of an incoming kindergartner. Children who meet or exceed the indicated score on all domains are identified as "kindergarten ready."[46]

The WaKIDS was piloted in a small number of elementary schools in the 2010-11 and 2011-12 AYs. Beginning in the 2012-13 AY WaKIDS administration was made a requirement for funding for full-day kindergarten.[47] Both full-day kindergarten and the WaKIDS were initially implemented in schools with the highest rates of children qualifying for free and reduced-price meals so that the population of children in earlier years is not representative of the full population of Washington kindergarteners. The number of schools implementing WaKIDS steadily increased through the 2017-18 AY, when WaKIDS was implemented in all public elementary schools.[48]

Because WSIPP's evaluation leverages multiple cohorts of students over time, changes that have been made to the assessments themselves during our study period have implications for WSIPP's evaluation. Although the conceptual definition of kindergarten readiness has remained consistent, even small changes in the measurement of an outcome can risk masking or misattributing effects of treatment on variation in that outcome. This is particularly problematic for methods intended to identify treatment effects on aggregate changes in an outcome over time.

In the 2015-16 AY, the specific set of objectives and dimensions included on the WaKIDS was revised based on WaKIDS data from earlier years. WaKIDS objectives/dimensions from 2015-16 through 2019-2020 have remained consistent. Because the 2014-15 AY WaKIDS was based on a different objective set, and we observed discontinuous breaks in the probability of kindergarten readiness on each domain from 2014-15 to 2015-16, we omitted the 2014-15 kindergarten cohort from all analyses.

[46] Education Research & Data Center. (2018). *Early learning feedback report.*
[47] RCW 28A.655.080.
[48] OSPI. *WaKIDS participation.*

Starting in the 2017-18 AY both ECEAP and kindergarten assessment data reflect a new version of the Teaching Strategies GOLD assessment. Specifically, Teaching Strategies introduced a version of the assessment designed to cover a developmental progression from birth through 3rd grade (B-3), whereas the previous version covered a developmental progression from birth through kindergarten (B-K).[49] Several simultaneous changes were made to the TSG B-3 version, including the addition of new objectives/dimensions on literacy and math domains, revising response scales to cover more advanced developmental knowledge and behaviors, revising some response scales to update and expand behavioral anchors (descriptions corresponding to objective scores), expanding the scale score range to accommodate the expanded developmental coverage, and finally, adjusting the benchmark scores for widely held expectations and kindergarten readiness based on these changes and new normative data using the TSG B-3 version. These changes all impact consistency of measurement over time for both the TSG and the WaKIDS.

To mitigate the impact of these changes on WSIPP's evaluation, we utilized objective/dimension level data to reconstruct scores for each domain that were as consistent as possible with the TSG B-K version. Our approach necessarily differed slightly for TSG and WaKIDS data.

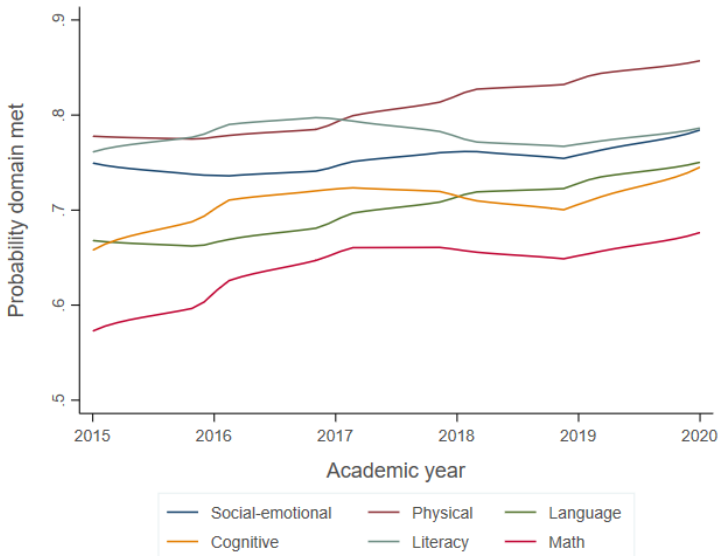
For the WaKIDS we summed objective/dimension scores for each domain. Teaching Strategies' approach in the TSG B-K version to missing objective level data was to impute the mean domain score for each child based on their completed items when at least 80% of items were completed. When fewer than 80% of items in a domain were complete mean imputation was not used, and the child's domain score was considered missing.[50] We replicated this approach and also checked that all students missing data for their kindergarten readiness flag in the original data we received from OSPI/ERDC were also missing this flag in our reconstructed data. We then applied WaKIDS benchmark cutoff scores as documented by Teaching Strategies.[51] No new items were added to the WaKIDS assessment when the B-3 version was introduced, and only the WaKIDS literacy raw score cutoff changed starting in 2017-18. This change was likely due to adjustments in the underlying response scales for dimensions in the literacy domain so that the B-K and B-3 cutoff versions were functionally equivalent. Based on inspection of the dimension response scales and data, we determined that the best approach to maintaining consistency over time in the kindergarten readiness classification was to apply the B-K version cutoffs through 2016-17, and the B-3 version cutoffs from 2017-18 forward. Exhibit A14 below depicts kindergarten readiness probabilities over time using this approach (Panel A), compared to the approach of applying the B-K version cutoffs in all years (Panel B).

[49] Lambert, R. (2017). *Technical Manual for the Teaching Strategies GOLD™ Assessment System: Birth through third grade rdition*.
[50] Lambert, R., Kim, D., & Burts, D. (2014). Technical manual (3rd edition) for the *Teaching Strategies GOLD* Assessment System.
[51] K. Houser, Teaching Strategies (personal communication, November 16, 2020).

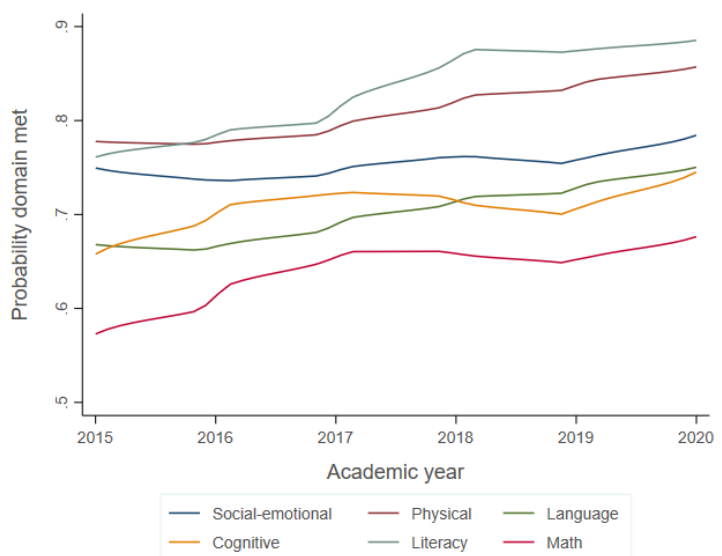
Exhibit A14

Probability of Kindergarten Readiness Over Time,
Comparing WaKIDS Raw Score Domain Cutoff Versions

Panel A



Panel B



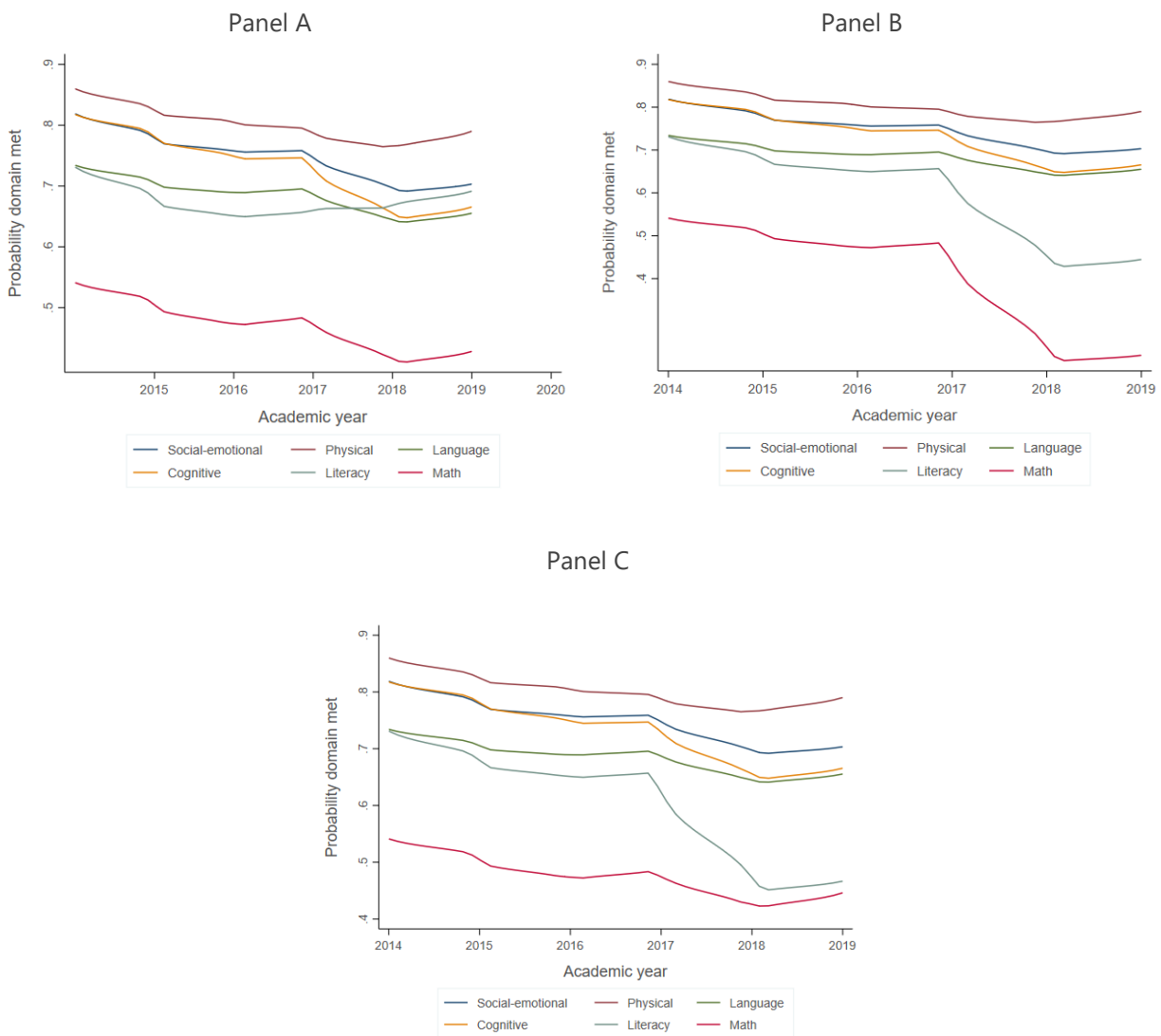
Note:
Figures show the unadjusted probability of meeting/exceeding the kindergarten readiness benchmark score in each domain.

For the TSG our approach was to (a) construct raw scores for each domain using only the items that were on the full TSG B-K version and then (b) apply the B-K version cutoffs to those raw scores for all years. To construct raw scores with a consistent item set over time we implemented the same missing data approach described above for the WaKIDS and summed scores within the domain. We chose to apply the B-K version cutoffs for all years because it was not possible to isolate changes in the benchmark cutoff

scores that were related to changed response scales (as opposed to new items or renorming). For the TSG, changes to the benchmark cutoff scores in the B-3 version were the result of several adjustments implemented all together, including both the addition of new items on some domains and revised and expanded response scales for many objectives/dimensions. We determined that applying the B-K cutoff scores produced the greatest consistency in measurement over time. Exhibit A15 below depicts kindergarten readiness probabilities over time using this approach (Panel A) compared with applying B-3 cutoffs to our reconstructed B-K raw scores starting in 2017-18 and (Panel B) to applying B-K and B-3 cutoffs to existing Full GOLD raw scores (Panel C).

Exhibit A15

Probability of Spring ECEAP Kindergarten Readiness Over Time,
Comparing Full GOLD Kindergarten Readiness Raw Score Domain Cutoff Versions



Note:
Figures show the unadjusted probability of meeting/exceeding the kindergarten readiness benchmark score in each domain.

V. Empirical Strategy

Ideally, we would evaluate the impact of site-level QRIS participation on child outcomes using a random assignment approach to first assign early learning sites to QRIS participation and then assign children to early learning sites. True random assignment allows us to unbiasedly attribute changes in outcomes across treatment and comparison groups to QRIS participation and not to unobservable (or observable) site-level *and* child-level confounding characteristics—examples of confounding characteristics include "baseline site quality" or "parenting style," respectively. However, because this is an observational study using administrative data, we are unable to use this approach, and in reality, sites (to some degree) select into QRIS participation and families choose which early learning sites to attend.

In order to mitigate bias due to child-selection into attending a site that has been rated and (conditional on rating) a site with a passing rating, we implement a statistical approach known as entropy balancing (weighting). Entropy balancing is a method that reweighs the comparison observations such that the mean and variance of selected control variables are the same in the treatment and comparison groups.[52]

Our entropy balanced estimation approach does not directly address potential bias due to site-level selection into treatment.[53] This is of particular concern when assessing the relationship between QRIS participation and child outcomes given that the choice to initiate the rating process (before a deadline) is left to the discretion of the early learning site. Before the ESA mandate (2015) QRIS-participation was voluntary, therefore early-adopters may be the most motivated sites (with greater resources) that produce more desirable child academic outcomes regardless of QRIS program completion. In order to control for level differences in child outcomes across sites, we combine our entropy matching method with a site-level fixed-effects approach for our analysis exploring research question one.

Entropy Balancing

Entropy balancing is a data preprocessing method that achieves balance on a set of pre-determined user-specified covariates in an observational study with a binary treatment variable.[54] This method directly estimates weights (rather than the propensity score) that solve a constrained optimization problem such that the reweighted treatment and comparison group balance on covariates incorporating information about known sample moments (e.g., mean, variance, skewness) and minimizing entropy distance (i.e., "uncertainty").[55] In other words, entropy weights allow us to exactly adjust for inequalities in observable predictors across the two groups (with regards to not only the mean but also higher moments of the predictor variable distribution).

[52] We perform entropy balancing using Stata's user-written program "ebalance" (Hainmueller & Xu (2013) and applying the default tolerance level of 0.015.
[53] Although there exist statistical methods that allow for matching at both the site-level and child-within-site-level, we do not have sufficient statistical power or variation in rating (level) status to confidently estimate reliable results using these techniques.
[54] In our study, the primary treatments indicators are "attending a rated site "and "attending a site rated at-quality."
[55] Hainmueller, J. (2012). Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Political analysis*, 25-46.

There are several proposed advantages of entropy weighting over alternative statistical data prepossessing methods (e.g., propensity score matching, exact matching, covariate balancing propensity score).[56] First, entropy balancing is more flexible than alternative data preprocessing methods that either match or discard observations—such as nearest neighbor or exact matching. Furthermore, in the case of exact matching, there is a trade-off between the number of matching covariates and the matched sample size (i.e., "the curse of dimensionality"). Entropy weighting reweights observations such that the maximum number of observations are preserved, even for large sets of covariates, without compromising balance—i.e., external validity is preserved without compromising internal validity. These benefits are particularly pronounced in the case of small sample size, with several low-probability covariates, and relatively unequal sample sizes across treatment and comparison groups—such as with our study.

Second, most methods (even those that similarly reweight observations on a continuum) do not directly focus on achieving balance on predictor variables. In practice, researchers often manually iterate across different matching covariates and propensity score models until a satisfactory balance is achieved. Furthermore, matching can offset reductions in bias when adjustments to propensity scores improve balance for some covariates while worsening balance for others.

In our analysis, we balance on a number of covariates that may affect the probability that a child attends a rated site and/or child outcomes.[57] This is the "matching step." Exhibit A16 and A17 show the difference-in-means balance test results across the treatment group (i.e., attending a rated site) and comparison group (i.e., attending an unrated site)—Column (1) summarizes the results for the unweighted sample, and Column (2) summarizes the same results for the entropy weighted sample. Exhibit A16 and A17 pertain to the sample of ECEAP attendees and subsidy attendees, respectively. As expected, results indicate that entropy weighting ensures exact balance on the means of these predictor variables.

[56] MacDonald, J.M., & Donnelly, E.A. (2019). Evaluating the role of race in sentencing: An entropy weighting analysis. *Justice Quarterly*, *36*(4), 656-681.

[57] Child outcomes include, probability of attaining a meet/exceed on at least 4/5/6 TS Gold and WA Kids domains, the probability of special education enrollment, and the average number of unexcused monthly absences.

Exhibit A16

Balance Test: ECEAP Child Characteristics, by Rated Status

Characteristics	Unmatched sample (1)	Matched sample (2)
Female	0.002	0.000
	(0.013)	(0.016)
Race:		
Black	0.030***	0.000
	(0.010)	(0.014)
Hispanic/Latinx	0.022	0.000
	(0.032)	(0.034)
Other	0.026***	0.000
	(0.009)	(0.012)
White	-0.078**	0.000
	(0.032)	(0.030)
Primary language:		
English	-0.027	0.000
	(0.026)	(0.028)
Spanish	0.010	0.000
	(0.025)	(0.028)
Other	0.016**	0.000
	(0.008)	(0.010)
Funding model:		
Part-day	-0.065***	0.000
	(0.019)	(0.031)
School-day	0.056***	0.000
	(0.018)	(0.028)
Work-day	0.008	0.000
	(0.007)	(0.012)
In ECEAP care in the previous year	0.144***	0.000
	(0.033)	(0.037)
Simultaneously in ECEAP care and kindergarten	-0.028***	0.000
	(0.006)	(0.004)
Full-time kindergarten	0.022	0.000
	(0.016)	(0.013)

Notes:
Each column and each row is derived from an alternative regression.
Estimates from Column (1) are derived from the entire unmatched sample.
Estimates from Column (2) are estimated from the entropy balanced sample.
***significant at the 1%-level, **significant at the 5%-level, and *significant at the 10%-level.

Exhibit A16, Continued

Balance Test: ECEAP Child Characteristics, by Rated Status

Characteristics	Unmatched sample (1)	Matched sample (2)
Mother married	-0.016	0.000
	(0.012)	(0.015)
Mother's education at birth:		
Less than high school	-0.019	0.000
	(0.020)	(0.023)
High school	0.004	0.000
	(0.012)	(0.014)
More than high school	0.015	0.000
	(0.015)	(0.020)
Mother's age at birth	0.512***	-0.001
	(0.162)	(0.189)
First-born	-0.027**	0.000
	(0.012)	(0.013)
Premature birth	-0.009	0.000
	(0.006)	(0.008)
Kindergarten characteristics		
Percent female	0.001	0.000
	(0.001)	(0.002)
Percent low-income	-0.011	0.000
	(0.014)	(0.016)
Percent section 504	0.006***	0.000
	(0.001)	(0.001)
Percent diagnosed disabled	0.000	0.000
	(0.003)	(0.003)
Percent enrolled in kindergarten	-0.005	0.007
	(0.013)	(0.019)
Number of observations (child x year)	10,278	10,278

Notes:

Each column and each row is derived from an alternative regression.

Estimates from Column (1) are derived from the entire unmatched sample.

Estimates from Column (2) are estimated from the entropy balanced sample.

***significant at the 1%-level, **significant at the 5%-level, and *significant at the 10%-level.

Exhibit A17

Balance Test: Subsidy Child Characteristics, by Rated Status

Characteristic	Unmatched sample (1)	Entropy balance (2)
Female	-0.013	0.000
	(0.011)	(0.011)
Race:		
Black	0.009	0.000
	(0.010)	(0.011)
Hispanic/Latinx	-0.037**	0.000
	(0.018)	(0.017)
Other	0.017	0.000
	(0.011)	(0.012)
White	0.010	0.000
	(0.021)	(0.021)
Primary Language:		
English	0.036***	0.000
	(0.013)	(0.011)
Spanish	-0.040***	0.000
	(0.011)	(0.010)
Other	0.004	0.000
	(0.005)	(0.006)
Previously received subsidy care	0.071***	0.000
	(0.012)	(0.012)
Simultaneously in subsidy care and kindergarten	-0.098***	0.000
	(0.013)	(0.014)
Full-time kindergarten	0.014**	0.000
	(0.006)	(0.004)
Married	-0.012	0.000
	(0.011)	(0.010)
Mother's education at birth:		
Less than high school	-0.036***	0.000
	(0.011)	(0.011)
High school	-0.010	0.000
	(0.011)	(0.011)
More than high school	0.046***	0.000
	(0.012)	(0.013)
Mother's age at birth	0.238*	0.000
	(0.132)	(0.138)
First-born child	0.005	0.000
	(0.010)	(0.011)
Premature birth	0.006	0.000
	(0.006)	(0.007)

Notes:
Each column and each row is derived from an alternative regression.
Estimates from Column (1) are derived from the entire unmatched sample.
Estimates from Column (2) are estimated from the entropy balanced sample.
***significant at the 1%-level, **significant at the 5%-level, and *significant at the 10%-level.

Exhibit A17, Continued

Balance Test: Subsidy Child Characteristics, by Rated Status

Kindergarten characteristics	Unmatched sample (1)	Entropy balance (2)
Percent female	-0.001	0.000
	(0.001)	(0.001)
Percent low-income	-0.031***	0.000
	(0.010)	(0.011)
Percent section 504	0.002*	0.000
	(0.001)	(0.001)
Percent diagnosed disabled	0.003	0.000
	(0.002)	(0.002)
Percent enrolled K-3	-0.009	0.002
	(0.005)	(0.005)
Number of observations (child x year)	9,153	9,153

Notes:
Each column and each row is derived from an alternative regression.
Estimates from Column (1) are derived from the entire unmatched sample.
Estimates from Column (2) are estimated from the entropy balanced sample.
***significant at the 1%-level, **significant at the 5%-level, and *significant at the 10%-level.

Fixed-Effects Model

As mentioned previously, in addition to concerns regarding child-level selection into treated sites, we are concerned about bias resulting from site-level selection into treatment. That is, the propensity for a site to opt into completing rating may be determined by unobservable site characteristics that vary across sites and affect child outcomes (e.g., sites of higher baseline quality may be most likely to first opt into QRIS program participation, and worst-performing sites may be most likely to never complete the program). Exhibits A18 and A19 summarize information from difference-in-means balance test results across treatment status for site-level characteristics (for ECEAP and subsidy site samples, respectively). Column (1) indicates that several site and neighborhood (i.e., census tract) characteristics vary significantly for children who attend a rated site versus children that do not. Furthermore, the results in Column (2) indicate that even when we apply the entropy balance weights on child-level characteristics, significant differences persist across site-level characteristics.

To lessen bias stemming from site-level selection, we estimate a site-level fixed-effects model. The fixed-effects model estimates site-specific intercepts that capture heterogeneities (i.e., level differences in outcomes) across sites.[58] This effectively controls for observed and unobserved time-invariant ("fixed") differences across sites that both predict rating status and affect child outcomes. For example, site fixed effects would account for time-invariant unobserved site-director attitude/motivation. This estimation strategy ultimately addresses concerns over selection bias because it removes all variation between sites from our treatment-parameter estimation. Therefore, the fixed effects model allows us to estimate changes in student outcomes that correspond to changes in treatment status *within-sites* as opposed to differences in treatment status *across-sites*. In our study, within-site variation in treatment status stems from changes in rating status across time (i.e., we exploit *temporal variation* in treatment); thus, we additionally control for pre-k academic year fixed effects. The inclusion of academic year fixed effects allows us to control for annual shocks that independently affect outcomes and are shared across sites

[58] In practice, this is achieved by including a separate dummy variable indicating each site in the regression model.

(e.g., federal government policy or a national recession). In addition, we include time-varying site-level covariates in our specification because fixed effects will not account for time-varying heterogeneity at the site-level.[59]

Our final model for evaluating research question one combines two-way fixed effects with entropy balancing. The matching method will address bias due to child-level selection on observables, whereas site fixed effects address selection on time-invariant unobservables. We estimate our model with an OLS regression that controls for all child-level, kindergarten school-level, and time-varying site-level characteristics (summarized in Appendix III). Standard errors are estimated to account for within-site clustering.[60]

Within our sample, sites rarely change the rating level. Therefore, we cannot implement a site fixed effects estimation strategy to explore research question two.[61] Instead, we assess the relationship between rating level and child outcomes by exploiting cross-sectional variation in rating levels across sites.[62] For this analysis we additionally control for the following time-invariant site-level characteristics years-in-operation, Child Care Aware (CCA) region, an indicator for the initial rating (vs. re-rating or renewal rating), is a licensed child care site (for ECEAP), primary language ever non-English (for ECEAP), received coaching (for subsidy).[63]

[59] These controls include average monthly enrollment (subsidy); average annual enrollment (ECEAP); and the following annual census tract information: percent at least high school graduate, percent at least college graduate, percent at households below the poverty line, unemployment rate, median household income, population under the age of five, percent Black, percent Hispanic/Latinx, and percent white. Census tract variables utilize the American Community Survey Five-Year Estimates for 2014 through 2018.
[60]Analysis pertaining to WaKIDS outcomes additionally controls for a school-level indicator variable denoting years from first WaKIDS administration. This accounts for how many years a school had been administering the WaKIDS assessment, and the possibility that scores in a first year of administration may differ from later years as a result of experience. In addition, analysis on the sample of ECEAP children controls for the following child-level variables pulled from ELMS: Individualized Education Program (IE)P indicator, over income eligibility indicator, and household type (two-parent\one-parent\other) indicators. Last we control for CCA region-by-year fixed effects to allow sites in different regions of the state to follow different trajectories and account for differential shocks by region over time (this inclusion improves precision without practically changing coefficient estimates).
[61] With additional data in future studies there may be sufficient within site variation to exploit a within-site estimation strategy to answer research question two.
[62] For the question two analysis, we estimate logistic and Poisson models as an alternative to OLS regression models.
[63] Our empirical models for question 2 still control for the aforementioned child, school, and time-variant site characteristics as well as annual fixed effects. Standard errors are estimated to allow for clustering at the site-level. These models additionally control for pre-k site-level variable "years since observed rating level was received."

Exhibit A18

Balance Test: ECEAP Site Characteristics, by Rating Status

Characteristic	Unmatched sample (1)	Entropy balance (2)
Primary language ever non-English	0.036**	0.025
	(0.016)	(0.019)
ECEAP enrollment	-23.132	-20.576
	(21.051)	(19.078)
Is a licensed	0.101***	0.067*
	(0.029)	(0.035)
Years in operation	0.415**	0.343**
	(0.168)	(0.160)
Census tract covariates:		
% Pop 25 years and up – High school graduate or higher	-1.093	-0.970
	(1.131)	(1.152)
% Pop 25 years and up – Bachelor's degree or higher	1.417*	0.304
	(0.776)	(0.990)
Unemployment rate	-1.603***	-1.406**
	(0.491)	(0.443)
Log median household income	0.056**	0.032
	(0.024)	(0.026)
Percent families below the FPL	0.824	1.310
	(1.107)	(1.076)
Population under age five	2.483	-11.594
	(22.177)	(23.031)
Percent Black	1.251**	1.585***
	(0.498)	(0.435)
Percent Hispanic/Latinx	4.500*	3.855
	(2.551)	(3.364)
Percent white	-3.570**	-3.318**
	(1.395)	(1.374)
Number of observations	10,278	10,278

Notes:
Each column and each row is derived from an alternative regression.
Estimates from Column (1) are derived from the entire unmatched sample.
Estimates from Column (2) are estimated from the entropy balanced sample.
***significant at the 1%-level, **significant at the 5%-level, and *significant at the 10%-level.

Exhibit A19

Balance Test: Subsidy Site Characteristics by Rating Status

Characteristics	Unmatched sample (1)	Entropy balance (2)
Primary language ever non-English	-0.042***	-0.012
	(0.015)	(0.014)
Average monthly enrollment	1.474	0.256
	(3.236)	(3.285)
Years in operation	0.054	0.087
	(0.114)	(0.118)
Received coaching	0.199***	0.199***
	(0.036)	(0.036)
Child care center	0.073***	0.044***
	(0.015)	(0.013)
Census tract covariates:		
% Pop 25 years and up – Bachelor's degree or higher	2.885***	0.324
	(0.695)	(0.637)
% Pop 25 years and up – Bachelor's degree or higher	2.551***	0.676
	(0.781)	(0.826)
Unemployment rate	-1.087***	-0.747***
	(0.224)	(0.223)
Log median household income	0.058**	0.025
	(0.024)	(0.025)
Percent households below the FPL	-1.359	-0.158
	(0.943)	(0.940)
Population under age five	-42.294***	-14.871
	(13.237)	(13.106)
Percent Black	0.876	0.234
	(0.536)	(0.558)
Percent Hispanic/Latinx	-0.601	0.274
	(1.217)	(1.162)
Percent white	-0.276	-0.617
	(1.039)	(1.063)
Number of observations	9,153	9,153

Notes:
Each column and each row is derived from an alternative regression.
Estimates from Column (1) are derived from the entire unmatched sample.
Estimates from Column (2) are estimated from the entropy balanced sample.
***significant at the 1%-level, **significant at the 5%-level, and *significant at the 10%-level.

VI. Outcomes Analysis: Detailed Results

The following tables summarize the full set of estimated results from our primary outcomes analysis discussed in Section IV of the main report.

Exhibit A20

Pre-K Early Achievers Rated Status and Kindergarten Readiness

	At least 4 (1)	At least 5 (2)	All 6 (3)
Panel A: spring TS gold, ECEAP sites			
Rated	-0.009	0.001	-0.020
	(0.019)	(0.022)	(0.027)
Number of observations	10,278	10,278	10,278
Mean	0.896	0.816	0.617
Impact (%)	-1.032	0.103	-3.238
Standard deviation	0.305	0.387	0.486
Effect size	0.030	0.002	0.041
Panel B: WaKIDS, ECEAP sites			
Rated	0.038	0.038	0.040
	(0.028)	(0.030)	(0.026)
Number of observations	10,278	10,278	10,278
Mean	0.741	0.612	0.394
Impact (%)	5.074	6.293	10.072
Standard deviation	0.438	0.487	0.489
Effect size	0.086	0.079	0.081
Panel C: WaKIDS, subsidy sites			
Rated	-0.005	-0.003	-0.005
	(0.025)	(0.027)	(0.025)
Number of observations	9,153	9,153	9,153
Mean	0.688	0.566	0.382
Impact (%)	-0.713	-0.477	-1.421
Standard deviation	0.464	0.496	0.486
Effect size	0.011	0.005	0.011

Notes:
Each row and each column represent a separate entropy balanced weighted regression.
Marginal effects are reported.
Each model controls for the full set of control variables and adjusts standard errors for clustering at the site level. Where WaKIDS outcomes are used the model additionally controls for the first year of implementation in school.
***significant at the 1%-level, **significant at the 5%-level, and *significant at the 10%-level.

Exhibit A21

Pre-K Early Achievers Coaching Receipt and Kindergarten Readiness, Subsidy Sites

	At least 4 (1)	At least 5 (2)	All 6 (3)
Panel A: Received coaching			
Rated	0.015	0.003	-0.002
	(0.053)	(0.056)	(0.050)
Ever received coaching	-0.001	0.186	0.204
	(0.276)	(0.301)	(0.264)
Rated × ever received coaching	-0.030	-0.011	-0.004
	(0.056)	(0.059)	(0.052)
Number of observations	9,153	9,153	9,153
Mean	0.672	0.542	0.351
Standard deviation	0.470	0.498	0.477
Panel B: Hours of coaching per month			
Rated	0.005	0.040	0.053
	(0.044)	(0.045)	(0.040)
Average hours of coaching per month	0.203	-1.782	-2.256
	(2.635)	(2.972)	(2.652)
Rated × average hours of coaching per month	-0.003	-0.007	-0.007
	(0.007)	(0.007)	(0.006)
Number of observations	7,213	7,213	7,213
Mean	0.672	0.542	0.351
Standard deviation	0.470	0.498	0.477

Notes:
Each row and each column represent a separate entropy balanced weighted regression.
Marginal effects are reported.
Each model controls for the full set of control variables and adjusts standard errors for clustering at the site level. Where WaKIDS outcomes are used the model additionally controls for the first year of implementation in school.
***significant at the 1%-level, **significant at the 5%-level, and *significant at the 10%-level.

Exhibit A22

Pre-K Early Achievers Rated Status and Other Kindergarten Outcomes

	Special education enrollment (1)	Monthly unexcused absences (2)
Panel A: ECEAP sites		
Rated	-0.017	-0.182***
	(0.015)	(0.060)
Number of observations	10,278	10,278
Mean	0.117	1.202
Impact (%)	-14.542	-15.144
Standard deviation	0.321	1.073
Effect size	0.053	0.170
Panel B: subsidy sites		
Rated	-0.011	0.040
	(0.015)	(0.098)
Number of observations	9,153	9,153
Mean	0.079	1.003
Impact (%)	-13.858	3.999
Standard deviation	0.270	0.963
Effect size	0.041	0.042

Notes:

Each row and each column represent a separate entropy balanced weighted regression. Marginal effects are reported.

Each model controls for the full set of control variables and adjusts standard errors for clustering at the site level. Where WaKIDS outcomes are used the model additionally controls for the first year of implementation in school.

***significant at the 1%-level, **significant at the 5%-level, and *significant at the 10%-level.

Exhibit A23

Pre-K Rated At-Quality and Kindergarten Readiness, ECEAP sites

	At least 4 (1)	At least 5 (2)	All 6 (3)
Panel A: Spring TS Gold			
Rated at-quality	0.060**	0.069**	0.034
	(0.024)	(0.032)	(0.038)
Number of observations	7,844	7,844	7,844
Mean	0.851	0.784	0.570
Impact (%)	7.011	8.753	6.017
Standard deviation	0.356	0.412	0.495
Effect size	0.168	0.167	0.069
Panel B: WA Kids			
Rated at-quality	0.040	0.061**	-0.012
	(0.029)	(0.029)	(0.035)
Number of observations	7,844	7,844	7,844
Mean	0.685	0.548	0.370
Impact (%)	5.806	11.209	-3.346
Standard deviation	0.465	0.498	0.483
Effect size	0.086	0.124	0.026

Notes:

Each row and each column represent a separate entropy balanced weighted regression. Marginal effects are reported.

Each model controls for the full set of control variables and adjusts standard errors for clustering at the site level. Where WaKIDS outcomes are used the model additionally controls for the first year of implementation in school.

***significant at the 1%-level, **significant at the 5%-level, and *significant at the 10%-level.

Exhibit A24

Pre-K Rated At-Quality and Kindergarten Readiness, Subsidy Sites

	At least 4 (1)	At least 5 (2)	All (3)
Panel A: Rating level at-quality			
Rated at-quality	0.041**	0.031	0.033*
	(0.019)	(0.020)	(0.020)
Number of observations	4,380	4,380	4,380
Mean	0.660	0.539	0.340
Impact (%)	6.260	5.680	9.660
Standard deviation	0.474	0.499	0.474
Effect size	0.087	0.061	0.069
Panel B: Rating level greater than 3 (vs. 3)			
Rated higher than level 3	0.011	0.037*	0.022
	(0.020)	(0.022)	(0.022)
Number of observations	3,642	3,642	3,642
Mean	0.670	0.542	0.346
Impact (%)	1.666	6.752	6.389
Standard deviation	0.470	0.498	0.476
Effect size	0.024	0.074	0.046

Notes:
Each row and each column represent a separate entropy balanced weighted regression. Marginal effects are reported.
Each model controls for the full set of control variables and adjusts standard errors for clustering at the site level. Where WaKIDS outcomes are used the model additionally controls for the first year of implementation in school.
***significant at the 1%-level, **significant at the 5%-level, and *significant at the 10%-level.

Exhibit A25

Pre-K Rated At-Quality and Kindergarten Readiness, Subsidy Sites—Including the 2018-19 Pre-K AY

	At least 4 (1)	At least 5 (2)	All 6 (3)
Panel A			
Rated	-0.015	-0.014	-0.021
	(0.020)	(0.023)	(0.022)
Number of observations	11,763	11,763	11,763
Mean	0.676	0.547	0.359
Impact (%)	-2.195	-2.624	-5.783
Standard deviation	0.468	0.498	0.480
Effect size	0.030	0.028	0.042
Panel B			
Rated at-quality	0.017	0.003	0.026
	(0.017)	(0.019)	(0.022)
Number of observations	6,854	6,854	6,854
Mean	0.670	0.552	0.358
Impact (%)	2.608	0.486	7.262
Standard deviation	0.470	0.497	0.480
Effect size	0.037	0.005	0.054
Panel C			
Rated higher than level 3	0.009	0.023	0.025
	(0.016)	(0.018)	(0.017)
Number of observations	5,453	5,453	5,453
Mean	0.683	0.557	0.368
Impact (%)	1.327	4.173	6.844
Standard deviation	0.465	0.497	0.482
Effect size	0.019	0.047	0.052

Notes:
Each row and each column represent a separate entropy balanced weighted regression. Marginal effects are reported.
Each model controls for the full set of control variables and adjusts standard errors for clustering at the site level. Where WaKIDS outcomes are used the model additionally controls for the first year of implementation in school.
***significant at the 1%-level, **significant at the 5%-level, and *significant at the 10%-level.

Exhibit A26

Pre-K Rated At-Quality and Other Kindergarten Outcomes, ECEAP Sites

	Special education enrollment (1)	Monthly unexcused absences (2)
Rated at-quality	-0.036***	-0.268***
	(0.015)	(0.082)
Number of observations	7,827	7,844
Mean	0.160	1.403
Impact (%)	-22.203	-19.078
Standard deviation	0.367	1.319
Effect size	0.097	0.203

Notes:
Each row and each column represent a separate entropy balanced weighted regression.
Marginal effects are reported.
Each model controls for the full set of control variables and adjusts standard errors for clustering at the site level. Where WaKIDS outcomes are used the model additionally controls for the first year of implementation in school.
***significant at the 1%-level, **significant at the 5%-level, and *significant at the 10%-level.

Exhibit A27

Pre-K Rated At-Quality and Other Kindergarten Outcomes, Subsidy Sites

	Special education enrollment (1)	Monthly unexcused absences (2)
Panel A: At-quality		
Rated at-quality	-0.012	-0.026
	(0.013)	(0.050)
Number of observations	4,380	4,380
Mean	0.091	1.104
Impact (%)	-13.325	-2.375
Standard deviation	0.287	1.071
Effect size	0.042	0.024
Panel B: Rating level greater than 3 (vs. 3)		
Rated higher than level 3	-0.005	0.030
	(0.013)	(0.040)
Number of observations	3,625	3,642
Mean	0.086	1.094
Impact (%)	-5.408	2.713
Standard deviation	0.281	1.002
Effect size	0.017	0.030

Notes:
Each row and each column represent a separate entropy balanced weighted regression.
Marginal effects are reported.
Each model controls for the full set of control variables and adjusts standard errors for clustering at the site level. Where WaKIDS outcomes are used the model additionally controls for the first year of implementation in school.
***significant at the 1%-level, **significant at the 5%-level, and *significant at the 10%-level.

VII. Sensitivity and Limitations

We test the sensitivity of our estimates to alternative specifications, sample restrictions, and different matching procedures. The following subsections discuss the sensitivity of our results under these differing conditions. We end this section with a discussion about the empirical limitations of this study. For simplicity, our sensitivity analyses will focus on the following WaKIDS assessment outcomes (for both the ECEAP and subsidy samples): meet/exceed at least 4 domains, 5 domains, and all 6 domains.

Entropy Balancing

Alternative Model Specifications

We first estimate the association between attending a site rated at-quality and WaKids performance using multivariate logistic regressions on the full sample prior to applying the entropy balancing method (Panel A, Exhibits A28 and A29). The results from our preferred entropy weighted model (with the full set of control variables) are summarized in Panel C. A primary motivation for implementing a matching method to preprocess the data is that the standard regression model does not accurately adjust for differences in observed variables when the distribution of those variables between two groups are significantly different.[64] However, the results using the unweighted sample (Panel A) are generally substantively similar to the results from the entropy weighted sample (Panel C) for both the sample of ECEAP-attendees and subsidy-attendees. The notable exception being that the relationship between rating-at-quality and meet/exceed all six domains varies decreases drastically (switches sign) with weights suggesting evidence of selection into treatment with regards to this outcome.

We then estimate a regression for the entropy balanced sample that omits site-level control variables (Panel B, Exhibits A28 and A29).[65] Results generally indicate that the introduction of site-level control variables in the preferred model (moving from Panel B to C) attenuates results slightly, although the difference in estimates is not statistically significant.

Alternative Matching Method

In Panel D of Exhibit A29, we present the sensitivity of our estimates of the impact of attending a subsidy site rated "at-quality" on WaKids performance to an alternative matching method.[66] In particular, we use 1:1 nearest neighbor propensity score matching with replacement. Entropy weighting provides comparable estimates to propensity score matching but maximizes analysis sample size and maintains a consistent sample across varying outcomes (Columns 1-3).

[64] Rubin (2001).
[65] Recall that our model for "rated at-quality" omits site fixed effects and thus includes the both time-variant and time-invariant site-level covariates.
[66] The size of the comparison group for children attending an ECEAP site are is too small to conduct a similar matching exercise for this group.

Exhibit A28

Pre-K Rated At-Quality and Kindergarten Readiness, ECEAP Sites—Alternative Specifications

	At least 4 (1)	At least 5 (2)	All 6 (3)
Panel A: No entropy balancing			
Rated at-quality	0.063	0.056	0.054
	(0.040)	(0.038)	(0.044)
Number of observations	7,844	7,844	7,844
Mean	0.685	0.548	0.370
Impact (%)	9.130	10.266	14.587
Standard deviation	0.465	0.498	0.483
Effect size	0.135	0.113	0.112
Panel B: Entropy balance, no control variables			
Rated at-quality	0.070	0.093**	-0.017
	(0.051)	(0.045)	(0.050)
Number of observations	7,844	7,844	7,844
Mean	0.685	0.548	0.370
Impact (%)	10.243	16.995	-4.591
Standard deviation	0.465	0.498	0.483
Effect size	0.151	0.187	0.035
Panel C: Entropy balance, full controls			
Rated at-quality	0.040	0.061**	-0.012
	(0.029)	(0.029)	(0.035)
Number of observations	7,844	7,844	7,844
Mean	0.685	0.548	0.370
Impact (%)	5.806	11.209	-3.346
Standard deviation	0.465	0.498	0.483
Effect size	0.086	0.124	0.026

Notes:
Each row and each column represent a separate entropy balanced weighted regression.
Marginal effects are reported.
Each model controls for the full set of control variables and adjusts standard errors for clustering at the site level. Where WaKIDS outcomes are used the model additionally controls for the first year of implementation in school.
***significant at the 1%-level, **significant at the 5%-level, and *significant at the 10%-level.

Exhibit A29

Pre-K Rated At-Quality and Kindergarten Readiness, Subsidy Sites—Alternative Specifications

	At least 4 (1)	At least 5 (2)	All 6 (3)
Panel A: No entropy balancing			
Rated at-quality	0.017	0.029	0.019
	(0.020)	(0.021)	(0.021)
Number of observations	4,380	4,380	4,380
Mean	0.660	0.539	0.340
Impact (%)	2.526	5.399	5.442
Standard deviation	0.474	0.499	0.474
Effect size	0.035	0.058	0.039
Panel B: Entropy balance, no control variables			
Rated at-quality	0.049*	0.033	0.041
	(0.030)	(0.031)	(0.030)
Number of observations	4,380	4,380	4,380
Mean	0.660	0.539	0.340
Impact (%)	7.461	6.207	11.973
Standard deviation	0.474	0.499	0.474
Effect size	0.104	0.067	0.086
Panel C: Entropy balance, full controls (preferred specification)			
Rated at-quality	0.041**	0.031	0.033*
	(0.019)	(0.020)	(0.020)
Number of observations	4,380	4,380	4,380
Mean	0.660	0.539	0.340
Impact (%)	6.260	5.680	9.660
Standard deviation	0.474	0.499	0.474
Effect size	0.087	0.061	0.069
Panel D: Propensity score matching, full controls			
Rated at-quality	0.048	0.025	0.024
	(0.031)	(0.029)	(0.029)
Number of observations	5,770	4,620	2,996
Mean	0.889	0.779	0.566
Impact (%)	5.372	3.238	4.203
Standard deviation	0.314	0.415	0.496
Effect size	0.152	0.061	0.048

Notes:
Each row and each column in Panels A-C represents a separate entropy balanced weighted regression. Marginal effects are reported.
Each model controls for the full set of control variables and adjusts standard errors for clustering at the site level. Where WaKIDS outcomes are used the model additionally controls for the first year of implementation in school.
Estimates in Panel D come from data preprocessed using 1:1 nearest neighbor matching with replacement.
***significant at the 1%-level, **significant at the 5%-level, and *significant at the 10%-level.

Sample Restrictions

One concern may be that participation in the QRIS rating program may induce sorting behavior. That is, we are concerned children who will perform better regardless of site quality are most likely to move to sites that have already been rated and received a higher rating level. Exhibits A30 and A31 replicate our primary analysis omitting children that have moved from a previous site to the site observed in the current sample. These results (Panel A), in comparison to our baseline analysis, indicate no evidence of children sorting into rated sites for ECEAP sites. However, when we remove children who moved to rated subsidy care sites, we see substantial declines in estimated results, although results remain statistically insignificant, this may suggest sorting of higher-performing children into rated subsidy care sites. Results remain generally robust when we remove children who moved to an ECEAP or subsidy care site rated "at-quality" (Panel B).

Exhibit A30

Pre-K Rated At-Quality and Kindergarten Readiness, ECEAP Sites—
Omitting Children who Moved to Current Site

	At least 4 (1)	At least 5 (2)	All 6 (3)
Panel A: Rated			
Rated	0.035	0.038	0.042
	(0.027)	(0.029)	(0.028)
Number of observations	9,542	9,619	9,639
Mean	0.741	0.610	0.396
Impact (%)	4.719	6.167	10.699
Standard deviation	0.438	0.488	0.489
Effect size	0.080	0.077	0.087
Panel B: Rated at-quality			
Rated at-quality	0.044	0.052**	0.004
	(0.031)	(0.025)	(0.031)
Number of observations	7,453	7,453	7,453
Mean	0.690	0.574	0.382
Impact (%)	6.316	9.099	1.034
Standard deviation	0.463	0.494	0.486
Effect size	0.094	0.106	0.008

Notes:
Each row and each column represent a separate entropy balanced weighted regression.
Marginal effects are reported.
Each model controls for the full set of control variables and adjusts standard errors for clustering at the site level. Where WaKIDS outcomes are used the model additionally controls for the first year of implementation in school.
***significant at the 1%-level, **significant at the 5%-level, and *significant at the 10%-level.

Exhibit A31

Pre-K Rated At-Quality and Kindergarten Readiness, Subsidy Sites—
Omitting Children who Moved to Current Site

	At least 4 (1)	At least 5 (2)	All 6 (3)
Panel A: Rated			
Rated	-0.022	-0.047	-0.025
	(0.041)	(0.044)	(0.044)
Number of observations	3,533	3,533	3,533
Mean	0.698	0.574	0.368
Impact (%)	-3.208	-8.237	-6.850
Standard deviation	0.459	0.495	0.482
Effect size	0.049	0.096	0.052
Panel B: Rated at-quality			
Rated at-quality	0.006	0.024	0.072**
	(0.035)	(0.034)	(0.031)
Number of observations	1,815	1,815	1,815
Mean	0.687	0.561	0.338
Impact (%)	0.867	4.329	21.372
Standard deviation	0.464	0.496	0.473
Effect size	0.013	0.049	0.153

Notes:
Each row and each column represent a separate entropy balanced weighted regression. Marginal effects are reported.
Each model controls for the full set of control variables and adjusts standard errors for clustering at the site level. Where WaKIDS outcomes are used the model additionally controls for the first year of implementation in school.
***significant at the 1%-level, **significant at the 5%-level, and *significant at the 10%-level.

Our baseline results assess the impact of rating level regardless of whether the level corresponds to the initial rating or a subsequent rating (i.e., re-rating or renewal rating). In our analysis sample, 86% of ECEAP enrolled children and 80% of subsidy enrolled children are exposed to the initial rating. One concern may be that initial ratings are more informative about underlying quality than subsequent ratings.

Exhibits A29-A32 recreate our baseline analysis with the sub-sample of children who are exposed to the initial rating. Our results regarding the association between rating at-quality and child outcomes are largely robust to the initial rating restriction although less precisely estimated (this is expected with a smaller sample size).

Exhibit A32

Pre-K Rated At-Quality and Kindergarten Readiness, ECEAP Sites—Initial Rating

	At least 4 (1)	At least 5 (2)	All 6 (3)
Panel A: spring TS gold			
Rated at-quality	0.062**	0.110***	0.073*
	(0.025)	(0.033)	(0.042)
Number of observations	6,705	6,705	6,705
Mean	0.850	0.767	0.553
Impact (%)	7.307	14.394	13.298
Standard deviation	0.357	0.423	0.497
Effect size	0.174	0.261	0.148
Panel B: WA kids			
Rated at-quality	0.028	0.007	-0.008
	(0.033)	(0.029)	(0.033)
Number of observations	6,705	6,705	6,705
Mean	0.703	0.583	0.382
Impact (%)	4.014	1.155	-1.979
Standard deviation	0.457	0.493	0.486
Effect size	0.062	0.014	0.016

Notes:
Each row and each column represent a separate entropy balanced weighted regression.
Marginal effects are reported.
Each model controls for the full set of control variables and adjusts standard errors for clustering at the site level. Where WaKIDS outcomes are used the model additionally controls for the first year of implementation in school.
***significant at the 1%-level, **significant at the 5%-level, and *significant at the 10%-level.

Exhibit A33

Pre-K Rated At-Quality and Kindergarten Readiness, Subsidy Sites—Initial Rating

	At least 4 (1)	At least 5 (2)	All 6 (3)
Panel A: Rating level at-quality			
Rated at-quality	0.018	0.013	0.012
	(0.024)	(0.024)	(0.023)
Number of observations	3,389	3,389	3,389
Mean	0.652	0.530	0.333
Impact (%)	2.828	2.365	3.676
Standard deviation	0.476	0.499	0.471
Effect size	0.039	0.025	0.026
Panel B: Rating level greater than 3 (vs. 3)			
Rated higher than level 3	0.029	0.055**	0.027
	(0.020)	(0.024)	(0.023)
Number of observations	2,843	2,843	2,843
Mean	0.672	0.540	0.344
Impact (%)	4.278	10.164	7.729
Standard deviation	0.469	0.498	0.475
Effect size	0.061	0.110	0.056

Notes:
Each row and each column represent a separate entropy balanced weighted regression.
Marginal effects are reported.
Each model controls for site and year fixed effects, the full set of control variables, and adjusts standard errors for clustering at the site level. Where WaKIDS outcomes are used the model additionally controls for the first year of implementation in school.
***significant at the 1%-level, **significant at the 5%-level, and *significant at the 10%-level.

Fixed Effects Model

A general concern with these analyses is that our estimated differences in outcomes are driven by inherent differences across sites and sample composition.

More specifically, we are concerned that sites with higher (unobserved) quality at baseline were more likely to register for Early Achievers (particularly prior to the ESA) and to complete ratings and are thus overrepresented among rated sites. To explore this concern, we re-estimate our primary analysis regarding the relationship between QRIS program participation and kindergarten readiness omitting sites that completed the program before 2016 (prior to the passage of the ESA).[67] The results from this analysis are presented in Panel B of Exhibits A34 and A35. Results (comparing Panel A and Panel B) generally suggest that site-level selection into early adoption is negatively biasing our results. The results in Exhibit 34 suggest that the relationship between attending an ECEAP site that has been rated and kindergarten readiness strengthens with the omission of early adopters.

An additional concern may be that sites of inherently lower quality, or sites that lack the motivation/capability/resources to reach higher quality never complete a rating and exit this market. If this kind of attrition occurs it may bias our estimated results. In order to empirically assess this issue, we re-estimate our primary results restricting the sample to only those sites we observe in-operation throughout our entire sample period. The results from this analysis are summarized for ECEAP and subsidy sites in Panel C of Exhibits A34 and A35. Results from these analyses are virtually indistinguishable from our full-sample results.

[67] This cutoff is more salient for ECEAP sites, because subsidy sites still had four years after the passage of ESA to complete the rating process.

Exhibit A34

Pre-K EA Rated Status and Kindergarten Readiness, ECEAP Sites—
Alternative Site-Level Sample Restrictions

	At least 4 (1)	At least 5 (2)	All 6 (3)
Panel A: Full sample			
Rated	0.038	0.038	0.040
	(0.028)	(0.030)	(0.026)
Number of observations	10,278	10,278	10,278
Mean	0.741	0.612	0.394
Impact (%)	5.074	6.293	10.072
Standard deviation	0.438	0.487	0.489
Effect size	0.086	0.079	0.081
Panel A: No early			
Rated	0.078**	0.063*	0.053
	(0.037)	(0.038)	(0.038)
Number of observations	6,606	6,606	6,606
Mean	0.749	0.621	0.394
Impact (%)	10.414	10.145	13.452
Standard deviation	0.433	0.485	0.489
Effect size	0.180	0.130	0.108
Panel B: Balance			
Rated	0.049	0.055	0.048
	(0.037)	(0.038)	(0.032)
Number of observations	8,224	8,224	8,224
Mean	0.749	0.621	0.394
Impact (%)	6.542	11.340	9.816
Standard deviation	0.433	0.485	0.489
Effect size	0.113	0.114	0.098

Notes:
Each row and each column represent a separate entropy balanced weighted regression.
Marginal effects are reported.
Each model controls for site and year fixed effects, the full set of control variables, and adjusts standard errors for clustering at the site level. Where WaKIDS outcomes are used the model additionally controls for the first year of implementation in school.
***significant at the 1%-level, **significant at the 5%-level, and *significant at the 10%-level.

Exhibit A35

Pre-K EA Rated Status and Kindergarten Readiness, Subsidy Sites—
Alternative Site-Level Sample Restrictions

	At least 4 (1)	At least 5 (2)	All 6 (3)
Panel A: Full sample			
Rated	-0.005	-0.003	-0.005
	(0.025)	(0.027)	(0.025)
Number of observations	9,153	9,153	9,153
Mean	0.688	0.566	0.382
Impact (%)	-0.713	-0.477	-1.421
Standard deviation	0.464	0.496	0.486
Effect size	0.011	0.005	0.011
Panel A: No early			
Rated	-0.019	0.000	-0.010
	(0.028)	(0.031)	(0.030)
Number of observations	6,191	6,191	6,191
Mean	0.670	0.541	0.352
Impact (%)	-2.836	0	-2.841
Standard deviation	0.470	0.498	0.478
Effect size	0.041	0.001	0.021
Panel B: Balance			
Rated	-0.008	-0.003	0.000
	(0.025)	(0.027)	(0.025)
Number of observations	8,844	8,844	8,844
Mean	0.670	0.541	0.352
Impact (%)	-1.194	-0.555	0
Standard deviation	0.470	0.498	0.478
Effect size	0.017	0.006	0.001

Notes:
Each row and each column represent a separate entropy balanced weighted regression.
Marginal effects are reported.
Each model controls for site and year fixed effects, the full set of control variables, and adjusts standard errors for clustering at the site level. Where WaKIDS outcomes are used the model additionally controls for the first year of implementation in school.
***significant at the 1%-level, **significant at the 5%-level, and *significant at the 10%-level.

A final concern is that observed changes in outcomes are driven by coincident changes in related policies or circumstances (i.e., confounders)—for example, observed changes may result from coincident changes in staffing, other early learning policies, or economic conditions. In order to explore the potential for confounders, we regress relevant site characteristics on rating status. For instance, the results summarized in the first row of Exhibit A36 indicate the undergoing rating (i.e., treatment) has no significant relationship with the proportion of Black students enrolled in ECEAP and subsidy care sites.

Results largely suggest site-level characteristics or census tract characteristics do not systematically change with rating completion. For example, we see no response in licensing behavior and enrollment size which alleviates concerns that relevant policies related to expansion are driving (or masking) the estimated relationship between rating completion and child outcomes. A notable exception is that the census tract unemployment rate drops systematically with rating completion for subsidy care sites. This could indicate that economic conditions predict rating completion (at the census tract level), however, the magnitude of the difference is small.

Exhibit A36

Pre-K Early Achievers Rated Status and Site Characteristics

Characteristics	ECEAP (1)	Subsidy (2)
Site characteristics		
Proportion Black	0.008	0.006
Proportion Hispanic/Latinx	0.003	-0.033
Proportion other race	-0.015	0.002
Proportion white	0.004	0.026
Proportion of primary language English	-0.026	-0.034*
Proportion of primary language Spanish	0.027	0.023
Proportion of primary language other	-0.001	0.011
Proportion enrolled in kindergarten and daycare	-0.008	0.038
Proportion enrolled in fulltime kindergarten	-0.05	-0.014
Annual average enrollment	-1.926	
Is a licensed care center	0.017	
License capacity		-0.262
Proportion of enrollment subsidy		-0.006
Census tract characteristics:		
Unemployment rate	0.063	-0.468**
Log median household income	-0.013	0.006
Percent household below the FPL	-0.278	-0.366

Notes:
Each row and each column represent a separate entropy balanced weighted regression.
Marginal effects are reported.
Each model controls for site and year fixed effects and adjusts standard errors for clustering at the site level.
***significant at the 1%-level, **significant at the 5%-level, and *significant at the 10%-level.

Limitations

In sum, our results are robust to a plethora of alternative specifications, data preprocessing methods, and sample restrictions. This robustness lends assurance that, where significant, our predicted outcomes are related to site rating status and not alternate circumstances or site/child characteristics. However, we still caution a causal interpretation of the results.

In a two-way fixed effects model, the causal identifying assumption is that outcomes for children in treated sites would have evolved similarly to those who attended untreated sites in the absence of QRIS program participation. However, we do not have enough information from the pre-treatment periods to empirically assess the validity of this assumption. Furthermore, we cannot rigorously assess if results are being driven by confounding factors (e.g., change in the composition of staff coincident with EA).

With regards to child-level selection bias, in order to draw causal inferences from an entropy balancing method, we must assume that treatment assignment depends only on observed data, and there are no relevant unobserved differences between the treated and control groups (i.e., "selection on observables"). Critically, there is no way to test this reaching assertion.

VIII. Subgroup Analyses

Given program interest in addressing the kindergarten readiness gap, we examine whether associations between pre-k exposure to an Early Achievers rated site and child outcomes are comparable across groups. Due to insufficient underlying variation in rated vs. unrated site attendance and outcome values in our sample, our ability to conduct reliable subgroup analyses is extremely limited. Here we examine child race/ethnicity as a potential source of heterogeneity in associations between pre-k sites' rated status and kindergarten readiness.

These analyses and results are summarized in Exhibits A37 and A38. For the Hispanic/Latinx subsample, results indicate that attending a rated ECEAP or subsidy site largely associates with an increase in the probability of kindergarten readiness over all domain counts although estimates are too imprecise to discern statistical significance. Results indicate no relationship between site rating completion and kindergarten readiness across other racial groups.

We urge caution in the interpretation of these results due to the low variation in treatment status in each group. That is, the number of children attending an unrated pre-k site, relative to the number of children attending a rated pre-k site, is too small to estimate statistically reliable results.

Exhibit A37

Pre-K Early Achievers Rated Status and Kindergarten Readiness, ECEAP Sites, by Race/Ethnicity

	Black (1)	Hispanic/ Latinx (2)	White (3)	Other (4)
Panel A: At least 4 domains				
Rated	-0.117	0.029	-0.037	-0.020
	(0.127)	(0.052)	(0.044)	(0.085)
Number of observations	810	2,798	3,958	1,587
Mean	0.682	0.610	0.695	0.708
Impact (%)	-17.197	4.758	-5.322	-2.849
Standard deviation	0.466	0.488	0.461	0.455
Effect size	0.252	0.059	0.080	0.044
Panel B: At least 5 domains				
Rated	-0.139	0.050	-0.018	-0.032
	(0.117)	(0.053)	(0.049)	(0.081)
Number of observations	810	2,798	3,958	1,587
Mean	0.572	0.466	0.572	0.572
Impact (%)	-24.210	10.739	-3.142	-5.556
Standard deviation	0.495	0.499	0.495	0.495
Effect size	0.280	0.100	0.036	0.064
Panel B: All 6 domains				
Rated	-0.041	0.036	-0.022	-0.018
	(0.137)	(0.054)	(0.043)	(0.086)
Number of observations	810	2,798	3,958	1,587
Mean	0.375	0.290	0.370	0.389
Impact (%)	-10.881	12.450	-6.079	-4.521
Standard deviation	0.484	0.454	0.483	0.488
Effect size	0.084	0.080	0.047	0.036

Notes:
Each row and each column represent a separate entropy balanced weighted regression.
Marginal effects are reported.
Each model controls for site and year fixed effects. The full set of control variables and adjusts standard errors for clustering at the site level. Where WaKIDS outcomes are used the model additionally controls for the first year of implementation in school.
***significant at the 1%-level, **significant at the 5%-level, and *significant at the 10%-level.

Exhibit A38

Pre-K Early Achievers Rated Status and Kindergarten Readiness, Subsidy Sites, by Race/Ethnicity

	Black (1)	Hispanic/ Latinx (2)	White (3)	Other (4)
Panel A: At least 4 domains				
Rated	-0.148	0.025	-0.032	-0.029
	(0.125)	(0.052)	(0.044)	(0.086)
Number of observations	810	2,798	3,958	1,587
Mean	0.684	0.609	0.693	0.709
Impact (%)	-21.584	4.160	-4.673	-4.021
Standard deviation	0.465	0.488	0.461	0.454
Effect size	0.317	0.052	0.070	0.063
Panel B: At least 5 domains				
Rated	-0.176	0.049	-0.022	-0.042
	(0.116)	(0.053)	(0.050)	(0.080)
Number of observations	810	2,798	3,958	1,587
Mean	0.575	0.466	0.570	0.572
Impact (%)	-30.691	10.572	-3.794	-7.362
Standard deviation	0.495	0.499	0.495	0.495
Effect size	0.357	0.099	0.044	0.085
Panel B: All 6 domains				
Rated	-0.085	0.033	-0.023	-0.042
	(0.131)	(0.055)	(0.043)	(0.086)
Number of observations	810	2,798	3,958	1,587
Mean	0.374	0.288	0.368	0.392
Impact (%)	-22.816	11.589	-6.376	-10.691
Standard deviation	0.484	0.453	0.482	0.488
Effect size	0.176	0.074	0.049	0.086

Notes:
Each row and each column represent a separate entropy balanced weighted regression.
Marginal effects are reported.
Each model controls for site and year fixed effects. The full set of control variables and adjusts standard errors for clustering at the site level. Where WaKIDS outcomes are used the model additionally controls for the first year of implementation in school.
***significant at the 1%-level, **significant at the 5%-level, and *significant at the 10%-level.

In addition, we examine child pre-k attendance duration as a potential source of variation in the effects of attending a rated pre-k site. Specifically, we test whether results are consistent for children who attend one versus more than one year of pre-k. Results are summarized in Exhibits A39 and A40.

Overall, rated site attendance is related to an increased probability of kindergarten readiness for ECEAP-site attendees. However, our analysis suggests that the overall positive association between pre-k rated site attendance and meeting kindergarten readiness benchmarks is largely driven by the subsample of children who attend an ECEAP site for two years.[68]

[68] Note that we restrict the sample to children who attend an ECEAP site for more than one year, not children who attend a rated site for more than one year. However, for ECEAP sites overlap between the two groups is very high and thus statistically difficult to disentangle.

This result may indicate that we are additionally picking up the effect of confounding policies that went into effect around the same time that the majority of ECEAP sites completed rating (the 2016 AY), for example, policies that changed access/availability of ECEAP sites. Further exploration into this concern is needed to better interpret our findings. For children in pre-k subsidy sites, we found no discernable statistical differences for children with one versus more than one year of pre-k in the association of site rating status and kindergarten readiness.

Exhibit A39

Pre-K Early Achievers Rated Status and Kindergarten Readiness, ECEAP Sites, by Tenure in Care

	At least 4 (1)	at least 5 (2)	All 6 (3)
Panel A: Two years of ECEAP care			
Rated	0.042	0.067	0.171***
	(0.050)	(0.051)	(0.058)
Number of observations	3,810	3,810	3,810
Mean	0.763	0.631	0.405
Impact (%)	5.569	10.694	42.209
Standard deviation	0.425	0.483	0.491
Effect size	0.100	0.140	0.348
Panel B: One-year ECEAP care			
Rated	0.014	0.012	-0.011
	(0.028)	(0.031)	(0.028)
Number of observations	6,468	6,468	6,468
Mean	0.730	0.597	0.390
Impact (%)	1.899	2.038	-2.793
Standard deviation	0.444	0.491	0.488
Effect size	0.031	0.025	0.022

Notes:
Each row and each column represent a separate entropy balanced weighted regression.
Marginal effects are reported.
Each model controls for site and year fixed effects. The full set of control variables and adjusts standard errors for clustering at the site level. Where WaKIDS outcomes are used the model additionally controls for the first year of implementation in school.
***significant at the 1%-level, **significant at the 5%-level, and *significant at the 10%-level.

Exhibit A40

Pre-K Early Achievers Rated Status and Kindergarten Readiness, Subsidy Sites, by Tenure in Care

	At least 4 (1)	at least 5 (2)	All 6 (3)
Panel A: At least one-year subsidy care			
Rated	-0.009	-0.005	-0.029
	(0.039)	(0.041)	(0.038)
Number of observations	4,329	4,329	4,329
Mean	0.695	0.570	0.374
Impact (%)	-1.344	-0.861	-7.835
Standard deviation	0.460	0.495	0.484
Effect size	0.020	0.010	0.061
Panel B: One year or less subsidy care			
Rated	-0.013	-0.025	-0.007
	(0.034)	(0.037)	(0.038)
Number of observations	4,824	4,824	4,824
Mean	0.648	0.514	0.327
Impact (%)	-2.071	-4.877	-2.082
Standard deviation	0.478	0.500	0.469
Effect size	0.028	0.050	0.015

Notes:

Each row and each column represent a separate entropy balanced weighted regression.

Marginal effects are reported.

Each model controls for site and year fixed effects. The full set of control variables and adjusts standard errors for clustering at the site level. Where WaKIDS outcomes are used the model additionally controls for the first year of implementation in school.

***significant at the 1%-level, **significant at the 5%-level, and *significant at the 10%-level.

For further information, contact:
Rebecca Goodvin at 360.664.9077, Rebecca.Goodvin@wsipp.wa.gov

Document No. 20-12-2203