Evaluation of the Child Signature Program: Summary Report, 2012–15

Evaluation Office
2389 Gateway Oaks Drive, Suite 260
Sacramento, CA 95833
(916) 263-1050
www.ccfc.ca.gov

February 2017

Author Information

Robert Dean performed statistical analyses and wrote initial drafts of this report. Debra Silverman provided editorial review. Gretchen Williams assisted with writing of the Evaluation Summary. Richard Flores prepared the map of CSP sites. David Dodds, Deputy Director of the Evaluation Office, provided general direction. The contributions of additional staff at First 5 California and First 5 county commissions are gratefully recognized in the Acknowledgments section at the end of this report.

Suggested Citation

First 5 California. 2017. *Evaluation of the Child Signature Program: Summary Report, 2012–15.* Sacramento, CA: First 5 California.

Table of Contents

List of Tables

List of Figures

Child Signature Program Evaluation Summary

In July 2012, First 5 California (F5CA) launched the Child Signature Program (CSP) in partnership with county commissions, building upon the success of F5CA's Power of Preschool (PoP) program and other investments to expand quality to new classrooms and communities statewide. CSP was designed to serve children ages 0 to 5 at greatest risk of school failure by ensuring they attend high-quality early learning programs that promote children's healthy development and school readiness. The Commission committed an investment of $135 million over three years to increase the quality of early learning programs across the state.

Also in 2012, California was awarded a highly competitive Federal Race to the Top – Early Learning Challenge (RTT-ELC) Grant, requiring states to implement a quality rating and improvement system (QRIS). Perhaps the most notable outcome for the entire CSP program is that programs with CSP classrooms were well-positioned to participate in RTT-ELC, and in most cases, received a high-quality rating.[1]

About the Study

This report is the result of CSP analyses conducted using annual program data collected across all three program years, with special emphasis on data collected during the 2014–15 school year.

Evaluation Highlights

This report shows CSP was a success as evidenced by more than 72,000 children served, assessments of high quality for physical environment and teacher-child interactions, and improvement in child development. The evaluation shows classrooms were high quality, benefiting at-risk children and families. CSP supported quality in early learning programs for children where the educational divide is greatest.

- Over the CSP funding term, over 72,000 disadvantaged children were served, mainly in publicly funded classrooms that include need and income eligibility thresholds. The majority of classrooms were located in low-performing areas of the state, as defined by the then used Academic Performance Index (API)[2].

[1] The relationships between elements of California's QRIS, dimensions of quality, and child outcomes will be evaluated as part of Race to the Top – Early Learning Challenge. This report shows high-quality CSP classrooms had some effect on developmental outcomes for young at-risk children as measured by the Desired Results Developmental Profile (DRDP 2010) (see *Appendix D* of this report).

[2] API is a measure of student achievement in school catchment areas. For program development of CSP 1, areas with API scores in the bottom three deciles were defined as "low performing areas." API deciles are collected for all CSP sites and correspond to the API of the public school catchment area where the site is located.

- In 2014–15, 60 percent of children served in the 1,350 participating CSP classrooms were of Hispanic or Latino ethnicity. More than half were dual language learners (DLL) (58 percent), and Spanish speakers accounted for 70 percent of this DLL group.

- The number of infants and toddlers increased over the life of CSP from a low of 515 served to a high of 979 served.

- In addition to continuing to support preschool classrooms, future programs need to ensure quality support to classrooms serving infants and toddlers in both centers and family child care homes.

CSP classrooms implemented high quality practices,

- Teaching staff were well-qualified. By 2014–15, fifty-eight percent of teachers held a BA degree or higher, and one quarter of teachers held an AA degree. An estimated 59 percent of teaching staff in QE classrooms held ECE- or CD-related degrees, as opposed to 36 percent in MOE classrooms. The average number of pooled ECE or CD units held by teaching staff per classroom was higher for QE classrooms—71 compared to 54 units.

- Most CSP classrooms conducted developmental screenings using the Ages and Stages Questionnaire (ASQ); they used the Desired Results Developmental Profile for preschoolers (DRDP-PS) to better meet children's individual needs.

- Classroom quality also was evaluated by external assessors. Most classrooms demonstrated effective teacher-child interactions as measured by the Classroom Assessment Scoring System® (CLASS)—seventy-one percent of all classrooms met minimum quality standards. Ninety percent of classrooms received a global score of 5 or above on the Environment Rating Scale (ERS) assessment—an indicator of a high-quality environment.

- Early childhood educators reported the most important aspect of CSP was the Quality Essential Staff—these experts were essential for the total functionality and effectiveness of program improvements.

CSP classrooms supported children and families.

- Children demonstrated healthy development over the life of the program, across all school years, and across all DRDP-PS developmental domains.

- During the life of the program, parent participation varied. Two-thirds of children had at least one parent who participated in some form of parent engagement or support activity. Quality Essential Staff (QES) played a key role in communicating with and engaging families.

Program and Evaluation Design

As described in previous CSP evaluation reports for the 2012–13 and 2013–14 school years (First 5 California 2014; 2015), it is clear high-quality preschool leads to positive outcomes for children with high needs (EOP 2014, Duncan and Magnuson 2013). Longitudinal research shows high-quality preschool also can lead to positive outcomes much later in life (Campbell et al. 2014; Heckman and Masterov 2007; Schweinhart 2007). Additionally, cost-benefit analyses demonstrate investments in high-quality preschool generate substantial economic payoffs by reducing a range of social costs such as unemployment, drug or alcohol abuse, and crime (Rees, Chai and Anthony 2012; Schweinhart et al. 2005; Heckman and Masterov 2007; EOP 2014; Duncan and Magnuson 2013; Yoshikawa et al. 2013).

In California, a major obstacle experienced by underprivileged groups is access to high-quality Early Care and Education (ECE). In 2007, approximately half of California's disadvantaged and at-risk 3- and 4-year-olds did not attend preschool, and even fewer attended high-quality preschool (Karoly et al. 2007). To address the scarcity of high-quality ECE in California, F5CA allocated funding through CSP to improve the quality of ECE classrooms in low-performing areas throughout California. Long-term goals of the program were to narrow the achievement gap for at-risk children and improve lifetime academic achievement and associated life success for California's youngest children (see Appendix A: CSP Logic Model).

Eight counties (Los Angeles, Merced, San Diego, San Francisco, San Joaquin, Santa Clara, Ventura, and Yolo) participated in the CSP Request for Application (RFA) 1 (CSP 1) during the 2012–13 school year.[3] In 2013, classrooms from two additional counties, San Mateo and Orange, joined CSP through the CSP Request for Application 3 (CSP 3) after first completing an extensive readiness assessment under the CSP Request for Application 2 (CSP 2).

F5CA implemented CSP 1 with two classroom quality levels—Maintenance of Effort (MOE) and Quality Enhanced (QE) (First 5 California 2012a).[4] All classrooms were required to meet minimum quality criteria defined through RFAs 1 and 3. Staff from all CSP classrooms and sites were able to access the Early Education Effectiveness Exchange (E4), a professional development forum for sharing ECE best practices and information. In addition to these inputs, Quality Essential Staff supported QE classrooms by implementing three essential program elements: 1) instructional strategies and teacher-child interactions; 2) social-emotional development; and 3) parent engagement and support. QES included support provided by Program Coordinators (PC), Local Evaluators (LE), Early Education Experts (EEE), Family Support Specialists (FSS), and Mental Health Specialists (MHS). QES implemented

[3] This report uses data collected for CSP 1 and 3 classrooms and sites only—it does not cover classrooms or sites participating in CSP 2.
[4] CSP 3 does not include the MOE quality level. All classrooms participating under CSP 3 were QE classrooms.

program elements through activities such as teacher training; developmental screening and assessment; and parent outreach, support, and education.

Role of QES in CSP

The work of the QES in CSP was multifaceted and complex, and a full analysis of their work is well beyond the scope of this report. Teams of QES supported classrooms and clients in various ways throughout the program. Each were assigned specific roles and responsibilities. PCs and LEs generally worked to coordinate local and state work surrounding the implementation and evaluation of the program. PCs provided oversight and support to other QES, were responsible for compliance and meeting program requirements, and ensured classrooms had quality criteria in place. LEs, in particular, coordinated the bulk of data collection and associated research activities, and regularly interacted with F5CA to drive and shape the evaluation. Other QES were tasked specifically with implementing the three core program elements of CSP, and each was responsible for a specific element.

- EEEs were responsible for implementing instructional strategies, and supporting and enhancing teacher-child interactions.
- MHSs worked to enhance social-emotional development.
- FSSs worked to enhance parent involvement and to support families.

With support from PCs and LEs, the EEEs and FSSs were the two most essential QES, according to county reports and narratives gathered from teachers. Specifically, EEEs and FSSs provided services such as:

QES collaborated extensively to achieve program goals. Through collaboration, QES were able to use data to inform QI activities across the program…

- Classroom observations and data collection
- Reflective practice coaching, training, and support
- Coaching at the site level for administrators and other support staff
- Quality Improvement Plan development
- Technical assistance for various tools and resources (e.g., CLASS®, Teaching Pyramid, Strengthening Families)
- Workshops for parents and staff covering a variety of topics including child development, nutrition, transition to kindergarten, substance abuse, positive parenting, mathematics and numeracy, language and literacy in the home, smoking and tobacco cessation, sugary beverages, social-emotional development, and car seat safety

- Family engagement activities, including cultural fairs and celebrations, parent-teacher conferences, opportunities for parents to volunteer in the classroom, parent cafes, and kindergarten fairs

- Home visits

- Needs assessments (e.g., ASQ, ASQ-SE, Protective Factors)

- Developing family partnership agreements

QES implemented program elements through activities such as teacher training, developmental screening and assessment, and parent outreach and support. More generally, QES worked as advisors and strategists to identify and meet needs of clients; as systems experts designing useful procedures and processes to enhance quality and to help ECE staff become more effective in their work; as social workers to link clients to resources; and as coaches, mentors, and consultants. As originally designed, QES were to support only QE classrooms in CSP as part of their quality enhancement, but as the program developed, QES had much more influence throughout the program and beyond, not only across quality levels but also across system boundaries (i.e., school, home, and community), and throughout ECE. QES were *essential* for the total functionality and effectiveness of CSP.

QES collaborated extensively to achieve program goals. An important feature of CSP was the built-in collaborative process through which LEs, EEEs, FSSs, and MHSs assessed classrooms and children; collected, organized, and analyzed data to immediately improve classroom environments and to implement the core program elements. Through collaboration, QES were able to use data to inform QI activities across the program, such as coaching practices and relationships, classroom instructional practices and the quality of classroom interactions, physical classroom environments, family and school partnerships, family development plans, etc., and to drive continuous quality improvement (CQI) across the system.

Evaluation Design

As described in the program Logic Model (see Appendix A, page 38), the overarching evaluation question was: How well does CSP reduce the achievement gap for at-risk young children? Evaluation hypotheses were that quality enhancements, such as access to QES, increased parent engagement and outreach, increased developmental screening, enhanced classroom interactions, and enhanced classroom environments, would improve outcomes for at-risk children.

To help address the overarching evaluation question, eleven specific questions were outlined in Attachment B of CSP RFA 1 (First 5 California 2012a) as outcome and process questions. Data for these questions included process measures useful for examining how well CSP was implemented, how well it served the public and specific

target populations (i.e., children who are DLLs, children with special needs, and children of seasonal migrants), its cost effectiveness, and outcome measures of children's cognitive, social, and physical development. Outcome and process questions developed for the evaluation of CSP are discussed in Appendix B of this report. Evaluation challenges and lessons learned are summarized in Appendix C.

As stated previously, this report is the result of analyses of data collected across all three program years (2012–2015), with special emphasis on data collected during the 2014–15 school year. It covers classrooms participating in CSP 1 and CSP 3 only, and compares data from multiple school years, as appropriate. Additionally, the evaluation is quasi-experimental, comparing QE classrooms to a stratified random sample of MOE classrooms.

Table 1 shows the total number of QE and MOE classrooms, and Table 2 shows the three-year total of QE and MOE classrooms by county. Figure 1 shows the distribution of CSP 1 and 3 classrooms across California.

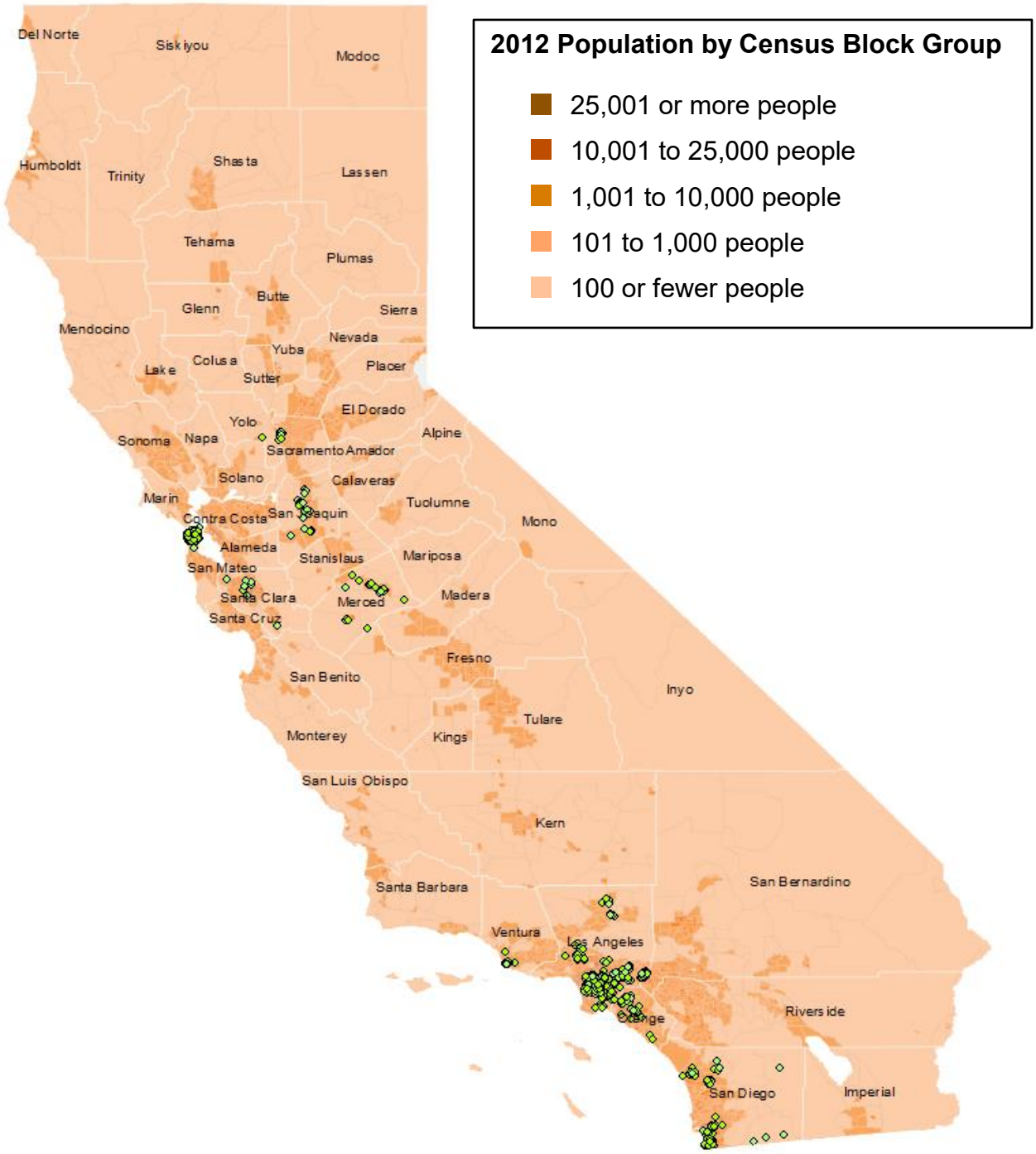
Table 1. Evaluation Design: 2014–15

Classroom Quality Level	Evaluation Classrooms	Non-Evaluation Classrooms	Total
QE	137	0	137
MOE	131	1,082	1,213
Total	263	1,082	1,350

Table 2. Classrooms by County and Classroom Quality Level: 2012–15

County	QE Classroom Records	MOE Classroom Records	Total CSP Classroom Records
Los Angeles	100	1,236	1,336
Merced	65	123	188
Orange	66	0	66
San Diego	42	940	982
San Francisco	23	919	942
San Joaquin	18	90	108
San Mateo	2	0	2
Santa Clara	18	128	146
Ventura	3	81	84
Yolo	34	52	86
All	371	3,569	3,940

Figure 1. CSP Site Locations and Population by Census Block Group



Note: Data Sources: CSP 1 and 3 site locations are from the CSP Profile and Evaluation Data system. Census block population estimates are from the Environmental Systems Research Institute (ESRI).

Characteristics of Participating CSP Programs and the Children and Families They Serve

CSP Programs

Most early learning programs in CSP are supported through multiple funding sources. In addition to funds to support quality improvement through F5CA, half of QE classrooms and more than one-third of MOE classrooms were funded through the California State Preschool Program (CSPP). Some classrooms also received state funding through the General Child Care Program (10 percent) and/or Alternative Payment Program (9 percent). In addition, 14 percent of classrooms in CSP received Head Start funding. Table D1 (pg. 47) shows the distribution of funding sources in 2014–15.

There was a marked decrease in the number of Head Start (404 fewer) and CSPP-funded (215 fewer) CSP classrooms between 2012–13 and 2014–15. This is possibly related to the federal sequestration in 2013 which was estimated to have impacted services for almost 4,000 California families, and upwards of 57,000 children across the nation (California Head Start Association, 2015). The 2011 recession also impacted the 2012–13 and 2013–14 California state budget allocations for state-funded preschool slots. Figure D1 (pg. 47) shows the change in numbers of classrooms funded for the top three funding sources for the 2014–15 school year.

Location of Classrooms by School Catchment Area

During development of CSP, F5CA defined areas with API scores in the bottom three deciles as “low-performing areas.” (See Table D2, pg. 48 and Figure D2, pg. 48). Fifty-two percent of classrooms were located in low-performing areas of the state in 2014–15. The initial distribution of classrooms across API in 2012–13 is likely explained by the “grandfathering” of classrooms from Power of Preschool (PoP) into CSP 1. A requirement of CSP is for classrooms to continue to serve at least 90 percent of children in the same targeted areas from PoP (First 5 California 2012a).

Children in CSP Classrooms

Children Served

CSP classrooms served over 23,500 children during the 2014–15 school year and over 72,000 children over the life of the program; the majority of these children were served in MOE classrooms. Table 3 (on the following page) shows most children (96 percent) in CSP were preschoolers (3- to 5-year-olds). The number of infants and toddlers served by CSP increased over time from a low of 515 served to a high of 979 served.

Table 3. Children Served by Age Group and Classroom Quality Level: 2014–15

	Preschoolers		Infants/Toddlers		Total Children Served	
	Number	Percent	Number	Percent	Number	Percent
QE	2,214	9%	216	1%	2,430	10%
MOE	20,447	86%	763	3%	21,245	90%
All Classrooms	22,661	96%	979	4%	23,640	100%

Note: percentages are for N = 23,640 children reported by age group and N = 23,675 total children served.

Special Target Populations

During the 2014–15 school year CSP served a total of 13,738 children who are DLL, 931 children with special needs, and 31 children from seasonal migrant families. Nearly two-thirds of all children in CSP were DLLs, and four percent of children were identified with special needs. QE classrooms served a higher proportion of DLLs than MOE classrooms (63 percent compared to 57 percent), but the same proportion (4 percent) of children with special needs.

QE classrooms served a slightly higher proportion of DLLs than MOE classrooms but the same proportion of children with special needs in 2014–15. Most DLLs spoke Spanish at home.

Race and Ethnicity

During the 2014–15 school year, CSP classrooms served a total of 13,400 children of Hispanic or Latino descent, which is 60 percent of total children served. Table 4 (on the following page) shows the distribution of ethnicity by MOE and QE classrooms. A greater proportion of children in QE classrooms (73 percent) were identified as Hispanic or Latino than were in MOE classrooms (59 percent). CSP was designed for children living in high needs areas, most often characterized by poverty. According to data from the National Center for Children in Poverty, QE classrooms served a representative proportion of Hispanic or Latino children living in poverty (NCCP 2014). Racial and ethnic service counts across all school years are shown in Figure D5 (pg. 51).

Table 4. Children Served by Race and Ethnicity and Classroom Quality Level: 2014–15

	QE		MOE		All	
	Number	Percent of Total Children Served	Number	Percent of Total Children Served	Number	Percent of Total Children Served
Hispanic/Latino	1,741	73%	11,659	59%	13,400	60%
Other	58	2%	1,901	10%	1,959	9%
White	172	7%	1,738	9%	1,910	9%
Asian	147	6%	1,970	10%	2,117	10%
Black or African American	97	4%	1,292	6%	1,389	6%
Two or More Races	77	3%	1,156	6%	1,233	6%
Native Hawaiian or Pacific Islander	15	1%	124	1%	139	1%
American Indian or Alaska Native	64	3%	59	<1%	123	1%
All	2,371	100%	19,899	100%	22,270	100%

Note: Percents are for N = 22,270 children reported by race or ethnic category among N = 23,675 total children served. Percent totals subject to rounding error.

Quality Indicators

Structural Quality

Researchers have identified three measurable structural program characteristics considered keys to structural quality: child-to-staff ratios, group size, and staff qualifications.

Most classrooms met CSP quality criteria for diversity, age groups, screening, classroom ratios, group size, staff qualifications, professional development, and use of ERS, CLASS®, and DRDP assessments.

Ratios and Group Size

All CSP classrooms met the required quality criteria based on Head Start, Title 5, and Title 22 requirements. More information about the ratios by quality level can be found in Tables D5a, D5b, and D6 (Pages 51 and 52).

Teacher/Provider Qualifications

CSP required lead classroom teachers and directors to hold a BA degree and 24 ECE units or meet the Child Development Permit Matrix (Permit) Program Director requirements. Assistant teachers in CSP were required to have an AA degree (or equivalent) with a minimum of 24 units of ECE. CSP teaching staff met or exceeded those requirements (see Figure D6, pg. 53); in 2014–15, more than half of all teaching staff held a BA (46 percent) or higher (6 percent). Other staff held an AA degree (24

percent) or completed some college (17 percent), with only 7 percent reporting a high school diploma as their highest level of education. In fact, there was a 14 percentage point increase in teachers with a BA degree between 2012–13 and 2014–15 (Figure D7, pg. 53). However, there were differences across quality levels as shown in Table 5, below. In 2014–15, a greater percentage of teachers in QE classrooms held AA, BA, or higher degrees (85 percent) than did their MOE peers (75 percent).

Table 5. Teaching Staff by Highest Level of Education and Classroom Quality Level: 2014–15

	QE		MOE		All Classrooms	
	Number	Percent	Number	Percent	Number	Percent
Less Than High School Diploma or GED	1	<1%	6	<1%	7	<1%
High School Diploma or GED	5	2%	127	8%	132	7%
Some College	27	12%	280	17%	307	17%
Associate's Degree	60	27%	391	24%	451	24%
Bachelor's Degree	117	53%	728	45%	845	46%
Advanced Degree	10	5%	93	6%	103	6%
Total	220	12%	1,625	88%	1,845	100%

Note: CSP teaching staff can work in multiple classrooms. Data used to create this table were collected as classroom-level data. Percents are based on N = 1,845 teaching staff records with data on highest level of education for approximately N = 1,162 teaching staff.

Table 5 shows a greater percentage of teachers in QE classrooms had degrees in ECE or CD than did teachers in MOE classrooms (59 percent of teachers in QE classrooms compared to 36 percent of teachers in MOE classrooms). By extension, teachers in QE classrooms had more ECE or CD units than their MOE peers and the difference is statistically significant (p<.0001).

Professional Development

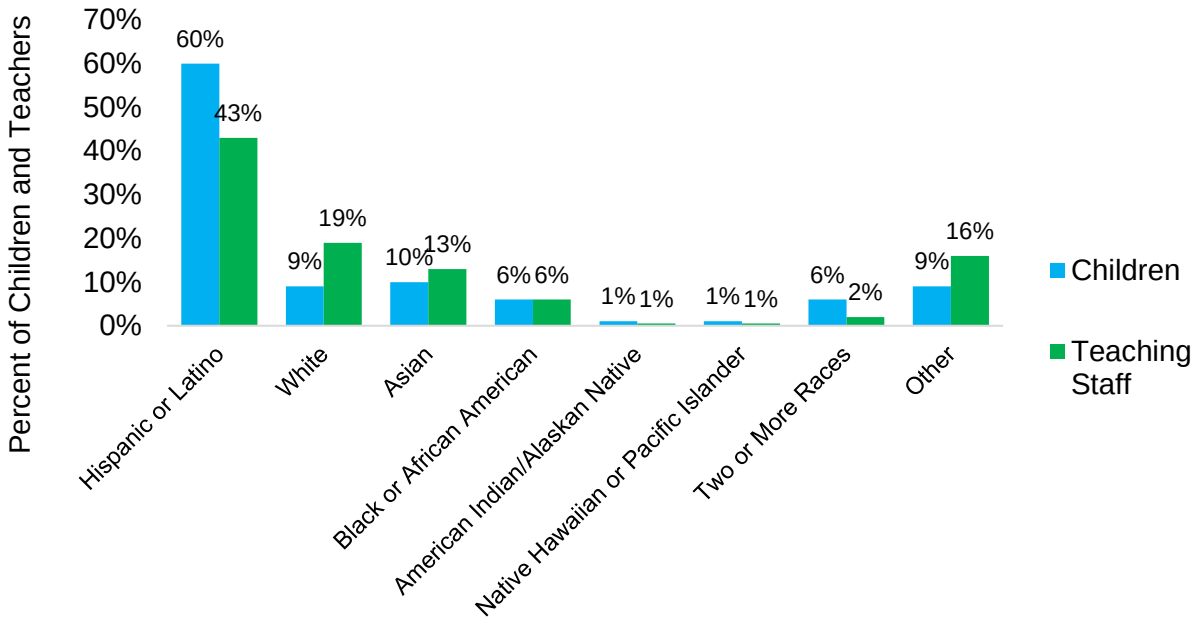
As a result of CSP, teachers participated in more professional development. There was a 56 percent increase in professional development activity between 2012–13 and 2014–15. (See Figure D10, pg. 55.)

Match Between Teacher and Child Ethnicity and Language

In terms of race and ethnicity, CSP classroom teaching staff were generally similar to the children they served during the 2014–15 school year. Figures D11 (pg. 55) and D12 (pg. 56) provide racial and ethnic information for teaching staff in CSP.

Figure 2 shows more children than teachers were identified as Hispanic or Latino (60 percent compared to 43 percent respectively) and more teachers than children were identified as White (19 percent compared to 9 percent, respectively).

Figure 2. Comparison of Teachers and Children by Racial and Ethnic Category: 2014–15



Note: Percents are based on an approximate N = 1,162 teaching staff and N = 22,269 children.

Environment and Interactions

Classrooms in CSP were evaluated by an external assessor using two tools: The Environment Rating Scales ERS and the CLASS®.

ERS tools are designed to assess the quality of ECE environments by helping assessors rate activities of children, teachers, other staff, and parents and their interactions with and within the environment (Cryer, Harms, and Riley 2003). CSP made use of three different ERS instruments to measure the quality of ECE environments: Early Childhood Environment Rating Scale (ECERS) is appropriate for children from 2 to 5 years old; Infant Toddler Environment Rating Scale (ITERS) is appropriate for children from birth to 2 years and 6 months old; and Family Child Care Environment

Rating Scale (FCCERS) is appropriate for FCC homes (Cryer, Harms, and Riley 2003, Harms, Clifford, and Cryer 2005; 2005a; 2005b).

CLASS® tools are designed to measure classroom quality by scoring interactions between children and teachers in classrooms as well as the teachers' use of the classroom environment (Pianta, Paro and Hamre 2008). CLASS Pre-K is appropriate for preschool classrooms serving children from 36 months of age to kindergarten entry (generally age 5), and CLASS Toddler is appropriate for classrooms serving toddlers between 18–36 months of age. CLASS Pre-K measures three domains of quality: Emotional Support, Classroom Organization, and Instructional Support; CLASS Toddler measures two domains: Emotional and Behavioral Support, and Engaged Support for Learning (Pianta, Paro, and Hamre 2008).

The majority of preschool classrooms met CLASS® Pre-K standards of 5 for Emotional Support, 3 for Classroom Organization, and 2.75 for Instructional Support. These results indicate teachers, on average, are using effective classroom interaction strategies as defined by CLASS

Environment Rating Scales

CSP required that classrooms meet and maintain a global score of 5 on the appropriate ERS tool indicating a 'good' level of quality as defined by the authors of the instruments (Harms, Clifford, and Cryer 2005). Table 6 shows more MOE classrooms met this standard than QE classrooms in 2014–15 . This pattern is true across the three years, as well. (See Table D10, pg. 56, Figures D13 and D14, pg. 57 and 58.)

Table 6. Distribution of Classrooms Meeting ERS Global Score Standards: 2014–15

	QE		MOE		All Classrooms	
	≥5	N	≥5	N	≥5	N
ECERS	80%	110	93%	395	90%	505
ITERS	76%	17	87%	15	81%	32
FCCERS	50%	4	67%	21	64%	25

While analysis did not detect statistically significant differences in global ERS scores between QE and MOE classrooms assessed with any ERS tool, there were some differences noted between QE and MOE classrooms when ECERS subscales were analyzed. In 2014–15, Space and Furnishings, Activities, and Parents and Staff subscale scores for MOE classrooms were significantly higher than those of QE classrooms. Interaction subscale scores were significantly higher in QE classrooms. (See Table D11, pg. 58.)

Classroom Assessment Scoring System®

External evaluators assessed the quality of classroom interactions with the CLASS Pre-K and CLASS Toddler instruments. CSP required classrooms to meet CLASS Pre-K domain scores of 5 for Emotional Support, 3 for Classroom Organization, and 2.75 for Instructional Support. Table 7 shows nearly all QE and MOE classrooms met the minimum quality standards in the Emotional Support and Classroom Organization domains, but only two thirds of QE and three-quarters of MOE classrooms met the minimum quality standards for the Instructional Support domain.

Table 7. Percentages of Observed Classrooms Meeting CLASS Pre-K Domain Standards: 2014–15

Domain (Domain score standard)	QE (N = 116)	MOE (N = 149)	Total (N = 265)
Emotional Support (≥5)	97%	99%	98%
Classroom Organization (≥3)	100%	100%	100%
Instructional Support (≥2.75)	66%	75%	71%
All Domains[a]	66%	75%	71%

Note: results are for all classrooms receiving a CLASS observation in 2014–15.
a. The Instructional Support standard seems to be a determining factor in whether or not a classroom met all CLASS Pre-K domain standards.

Statistical tests were unable to detect consistent differences in CLASS Pre-K domain scores over the life of the program. (See Tables D17, pg. 62 and D18 pg. 62.) Further, there was not a statistical difference between Toddler CLASS scores in QE and MOE classrooms in 2014–15 (Table D19, pg. 63).

There is some evidence indicating classrooms score higher on the Pre-K CLASS Instructional Support Domain when teachers have completed more ECE or CD units, regardless of whether they are teaching in QE or MOE classrooms. (See Tables D20 and D21, pg. 64.)

Parent Engagement and Support

Parents participated in a variety of parent engagement and support activities, such as advisory boards, parent teacher conferences, classroom volunteering opportunities, education to support parenting and child development, and other social support activities. Over the life of the program, parent participation by engagement and support activity was inconsistent. There was some variability in parent participation by type and over time. (See Table D22, pg. 65.) Parent-teacher conferences drew the most parent participation across all three years: in 2014–15, two-thirds of active parents participated in a parent-teacher conference (67 percent); this is an increase over 2013–14 (59 percent), but a decrease from 2012–13 (79 percent.) (See Figure D15, pg. 66.)

The majority of parents participated in parent-teacher conferences. However, parent participation in educational opportunities, volunteering, and social support activities was low.

These results should be interpreted cautiously since the number of total active parents fluctuated greatly over the life of the program, from 21,303 active parents in 2012–13 to a high of 31,823 active parents in 2013–14, and back to 15,771 active parents in 2014–15. There was an increase in the number of parents participating in educational opportunities (a low of 10 percent to a high of 16 percent). Percentages of active parents participating in classroom volunteer and social support activities also increased during the 2013–14 school year from 5 to 10 percent, and from 4 to 13 percent, respectively, but decreased to 8 and 6 percent in 2014–15.

These dramatic shifts in parent participation rates may be related to the 2013 federal sequestration and the drop in Head Start-funded classrooms because Head Start incorporates a strong family engagement component through its *Parent, Family, and Community Engagement Framework*. (See Head Start 2011.) However, while it is reasonable to suggest the drop in Head Start-funded classrooms also could have impacted family engagement activities at CSP sites, this factor alone probably does not account for all fluctuations in parent participation.

Child Development and School Readiness

CSP classrooms assessed child development and school readiness, and screened children for early intervention using the following tools:

- Ages & Stages Questionnaires®, Third Edition (ASQ-3™) measures developmental progress in children between the ages of 1 month to 5 ½ years and is used to identify delays or problems that can be referred for further assessment and specialized intervention. The ASQ Social-Emotional (ASQ-SE™) is focused entirely on social and emotional development.

- Desired Results Developmental Profile 2010 (DRDP 2010), including DRDP-PS, DRDP-IT, and DRDP *access*, is the primary measure of child development for

California Department of Education's (CDE) Desired Results (DR) system (CDE 2010; CDE 2011). DRDP instruments are authentic observational assessments, based on naturalistic and participant observation methodology, designed to guide teachers through the process of observing and documenting the development of children across a developmental continuum (McLean, Edelman, and Salcedo 2011, CDE 2010). DRDP 2010 includes three consecutive, yet overlapping, assessments corresponding to three age groups: DRDP Infant/Toddler (DRDP-IT) is appropriate for assessing children from birth to 36 months, and DRDP-PS is appropriate for assessing children from three years to kindergarten entry.[5] DRDP is both formative and summative. As a formative assessment, DRDP produces results most useful for informing classroom instruction, interaction, and processes at the classroom level. For this evaluation, DRDP-PS and DRDP-IT were used as summative assessments to measure how children developed in CSP classrooms at a programmatic level.[6] Teachers assess the development of each child by observing and documenting specific evidence demonstrating the child has mastered a particular developmental level in terms of a specific measure. Each child demonstrates where they are developmentally along the continuum for every measure under each developmental domain. This collection of ratings constitutes their developmental profile. Individual children's progress through the developmental levels are not analyzed for this report. Rather, DRDP profiles are aggregated to produce a distribution of DRDP ratings for the classroom.

Figure D16 (pg. 66) shows the percentages of CSP classrooms using a DRDP instrument and ASQ.

School Readiness

The main research focus was to show how well CSP reduced the achievement gap for young children, and whether greater gains were made by children in QE classrooms than in MOE classrooms. DRDP ratings data may be analyzed with various techniques. In the following narrative, ratings for fall and spring are tested for changes in proportions. Appendix D (tables D25, D26 and D27, pages 70 through 74), contains additional analyses for DRDP data that assesses shifts in ordinal rankings.

DRDP results should be interpreted with some caution for several reasons. First, teachers in CSP are typically not trained assessors. Teachers as observers are personally invested in the development of the children in their classrooms and may possibly inflate or deflate DRDP ratings for various reasons. Second, teachers possess various levels of understanding of the DRDP assessment instrument and procedures so child development may not be assessed the same way or with the same attention to detail across all CSP classrooms. DRDP data may be less reliable and consistent than

[5] Desired Results Developmental Profile-School Age (DRDP-SA) was not used in CSP classrooms.
[6] Formative means that assessment results are used to shape classroom instruction. The goal is to monitor child progress as feedback to inform classroom instruction. Summative implies outcomes to measure child development for purposes of comparison (i.e., to compare to some standard of development or to the development of some other group of children).

other assessment methods using independent observers and other child development assessment instruments. (Analyses of these data using the Cliff's Delta statistic may reduce these two possible biases [see *Appendix D, pages 68 through 74*]). Third, turnover of teaching staff in CSP classrooms may have influenced results if calculated by different teaching staff with different levels of understanding of DRDP.

Results are mixed. Table 8 compares percentages of ratings in the top two developmental levels during fall and spring across all developmental domains of DRDP-PS (CDE 2010a) by classroom quality level for the 2014–15 school year.[7] While fewer ratings from children in MOE classrooms were in the top two developmental levels during both the fall and spring assessments than QE classrooms, gains in QE and MOE were the same or nearly the same in four of the seven developmental domains from fall to spring (i.e., pre- and post-tests). Children in QE classrooms made greater gains in the Language and Literacy Domain, and children in MOE classrooms made greater gains in Physical Development and English Language. For English Language Development, slightly more than two-thirds of ratings were in the top two developmental levels at the end of the year.

[7] Percentages do not reflect percentages of children, but rather percentages of DRDP ratings. Children are rated across multiple measures and multiple dimensions when they are assessed using DRDP. The activities of one child will generate ratings at different developmental levels across multiple measures of multiple DRDP dimensions. Since the unit of analyses for the evaluation of CSP is the classroom and not individual children, the development of children is best understood as a constellation of DRDP ratings. The aggregate DRDP data collected does not differentiate between individual children, but rather utilizes the collective ratings of the children in the classroom in order to develop a developmental distribution of ratings for the classroom.

Table 8. Percentages of Ratings at the Top Two DRDP-PS Developmental Levels at Fall and Spring by Classroom Quality Level: 2014–15

		Percent Ratings at Top Two Developmental Levels			Difference in Percents (QE – MOE)		*N* Ratings	
Developmental Domain	Classroom Type	Fall	Spring	Gain	Fall	Spring	Fall	Spring
Self and Social Development	QE	37%	82%	+45%	5%***	5%***	15,159	14,287
	MOE	32%	77%	+45%			15,398	15,400
Language and Literacy Development	QE	26%	74%	+48%	1%*	5%***	12,662	11,568
	MOE	25%	69%	+43%			12,943	12,881
English Language Development	QE	42%	70%	+28%	14%***	1%	3,506	3,565
	MOE	28%	69%	+41%			3,463	3,594
Cognitive Development	QE	35%	81%	+46%	6%***	6%***	6,335	5,781
	MOE	29%	75%	+46%			6,408	6,419
Mathematical Development	QE	28%	77%	+49%	4%***	6%***	7,576	7,002
	MOE	25%	72%	+47%			7,661	7,698
Physical Development	QE	62%	94%	+32%	10%***	6%***	3,780	3,490
	MOE	52%	88%	+36%			3,859	3,851
Health	QE	43%	86%	+43%	5%***	6%***	3,784	3,435
	MOE	39%	80%	+41%			3,867	3,783

Note: N = number of *ratings*, not children. Some DRDP dimensions have more ratings because those dimensions have more measures.
* $p<.05$, ** $p<.01$, *** $p<.001$. Inconsistencies in differences in percentages are due to rounding.

Similar information is presented for all school years in Table D23 (pg. 67). For these combined data, there are statistically significant differences between percentages of ratings at the top two developmental levels for all seven developmental domains in both fall and spring. Children in QE classroom started higher than their counterparts in MOE classrooms and made greater gains. These combined data suggest, over the life of the program, QE classrooms were able to move higher percentages of ratings into the higher developmental levels than were MOE classrooms.

DRDP-IT data is shown in Table 9, below. With the exception of Motor and Perceptual Development, QE and MOE classrooms were not statistically different, though the overall pattern of change in the differences suggests infants and toddlers in MOE classrooms developed more than in QE classrooms. The most notable result was for Motor and Perceptual Development where MOE classrooms started with less, but ended with more, ratings in the top two developmental levels of DRDP-IT than QE classrooms—QE classrooms ratings in the top two developmental levels increased by

one percent, while MOE classroom ratings in the top two levels increased by 16 percent between fall and spring.

Table 9. Percentages of Ratings at the Top Two DRDP–IT Developmental Levels at Fall and Spring by Evaluation Classroom Quality Level: 2014–15

		Percent Ratings At Top Two Developmental Levels			Difference in Percents (QE – MOE)		*N* Ratings	
Developmental Domain	Classroom Type	Fall	Spring	Gain	Fall	Spring	Fall	Spring
Self and Social Development	QE	38%	47%	+9%	<-1%	-3%	1,971	1,581
	MOE	39%	50%	+11%			716	663
Language and Literacy Development	QE	21%	29%	+8%	-7%*	-8%*	908	723
	MOE	28%	37%	+9%			330	306
Cognitive Development	QE	40%	49%	+8%	3%	1%	1,616	1,305
	MOE	36%	48%	+12%			612	560
Motor and Perceptual Development	QE	52%	53%	+1%	4%	-10%*	599	480
	MOE	48%	64%	+16%			220	204
Health	QE	60%	62%	+2%	9%	-3%	146	118
	MOE	51%	65%	+14%			55	51

Note: N = number of *ratings*, not children. Some DRDP dimensions have more ratings because those dimensions have more measures.
* $p<.05$. Inconsistencies in differences in percentages are due to rounding.

Combined data from all three school years is shown in Table D24 (pg. 68). Results are mixed for these data, as well. Overall, toddlers in QE classrooms started lower and ended lower than their peers in MOE classrooms across all domains of learning. Children in QE classrooms made greater gains than children in MOE classrooms on the Self and Social Development, Cognitive Development, and Health domains. Children in MOE classrooms made greater gains than children in QE classrooms in the domains of Language and Literacy Development and Motor and Perceptual Development.

Summary and Conclusions

In summary, CSP classrooms are high-quality in terms of physical classroom environments and quality of interactions within those environments. Because of this, programs with CSP classrooms were well positioned to participate in RTT-ELC. In 2014–15, as in prior school years, CSP was able to reach underprivileged and low-income families and special target groups, classrooms were diverse in terms of race and ethnicity, teaching staff were well qualified, and children continued to demonstrate healthy development regardless of classroom quality level (QE or MOE).

CSP served over 23,500 children during the 2014–15 school year, and over 72,000 children during the life of the program. CSP served 13,738 DLL children, 931 children

with special needs, 979 infants and toddlers, and 13,400 children of Hispanic or Latino ethnicity—equating to 60 percent of total children served in 2014–15. Spanish-speaking DLLs accounted for 70 percent of all DLLs in 2014–15. Fifty-two percent of classrooms were located in school catchment areas in the bottom three deciles of API. The majority of CSP classrooms served children of families meeting either state or federal income eligibility standards (i.e., low-income).

Classroom teaching staff were well-qualified in 2014–15 and during the life of the program. Fifty-two percent of all teaching staff held at least a BA degree during the 2014–15 school year, and teaching staff with BA degrees increased to 46 percent (up from 32 percent in 2012–13). An estimated 59 percent of teaching staff in QE classrooms held ECE- or CD-related degrees, as opposed to 36 percent in MOE classrooms. The average number of pooled ECE or CD units held by teaching staff per classroom was higher for QE classrooms—71 units compared to 54 units. There was a 56 percent increase in teachers taking part in annual professional development provided through CSP sites over the life of the program.

QES collaborated extensively to make the program work. An important feature of CSP was the built-in collaborative process through which LEs, EEEs, FSSs, and MHSs assessed classrooms and children, and collected, organized, and analyzed data to improve classroom environments and implement the core program elements.

On average, classroom physical environments were above a “good” level of quality (i.e., ERS global score of 5 or above). Ninety percent of preschool classrooms, 81 percent of infant/toddler classrooms, and 64 percent of FCCs achieved ERS global scores of 5 or above in 2014–15. MOE classrooms showed higher levels of quality than did QE classrooms in 2014–15. Over the life of the program, statistical tests for differences in ECERS scores suggest MOE classrooms were able to catch up with, and then surpass, QE classrooms in terms of classroom quality. Ninety-eight percent of evaluation classrooms met CLASS® Emotional Support Domain score standards, 100 percent met Classroom Organization standards, and 71 percent met Instructional Support standards. Seventy-one percent of evaluation classrooms met all three CLASS domain standards.

In terms of child development, gains in percentages of ratings in the top two developmental levels of DRDP-PS by developmental domains for the combined three-year data suggest that, over the life of CSP, QE classrooms were able to move higher percentages of ratings into the higher DRDP-PS developmental levels. During 2014–15, consistent medium and large effect sizes across classroom quality levels for DRDP-PS suggest preschool children experienced healthy development, whether in QE or MOE classrooms. During the life of the program, consistent medium and large effect sizes across all DRDP-PS developmental domains, for both QE and MOE classrooms, suggest children showed healthy development whether in QE or MOE classrooms (Appendix D). In addition, teacher-reported DRDP-PS ratings on domains of child development across school years suggest QE preschool classrooms contributed more toward a reduction of the achievement gap for young children than MOE classrooms.

Successes

CSP was a success as evidenced by more than 72,000 children served, assessments of high quality for physical environment and teacher-child interactions, and improvements in child development. Though there were challenges in implementing a quasi-experimental design for CSP (Appendix C), the evaluation shows classrooms were high quality, benefiting at-risk children and families. As suggested by teachers' ratings with DRDP, children appear to have benefited especially from participation in QE classrooms (Appendix D). Collaboration between F5CA with local LEs and PCs enabled data collection and reporting, despite the complexities of local program and evaluation implementation. The *CSP Data Collection Guidebook* and monthly evaluation calls between F5CA and local evaluators allowed identification and discussion of processes for adapting evaluation design and data collection during the life of the program. Findings of the CSP evaluation support the utility of building and maintaining flexible, collaborative working relationships within an integrated ECE system to effectively serve California's most at-risk children and families.

Human Subjects Protection

Evaluation of CSP 1 and 3 was conducted under review of the State Committee for the Protection of Human Subjects, Protocol ID 12-08-0632.

Acknowledgments

F5CA acknowledges staff of the 10 First 5 county commissions, Los Angeles, Merced, Orange, San Diego, San Francisco, San Joaquin, San Mateo, Santa Clara, Ventura, and Yolo, that committed to CSP quality enhancement efforts. F5CA also acknowledges the local QES, the EEEs, FSSs, and MHS, as well as the CSP teaching staff, local program directors, and assessors who moved CSP from concept to reality.

Special thanks are made to CSP LEs, PCs, and other local staff who have worked closely with F5CA's Evaluation Office to help design, re-design, and realize evaluation goals.

Los Angeles
Adrienne Coppola, Los Angeles Universal Preschool (LAUP)
Kevin Dieterle, First 5 Los Angles
Donna Escalate, LAUP
Aleece Kelly, First 5 Los Angeles
Dawn Kurtz, LAUP
Laura Marin, LAUP
Matilde Morales, LAUP
Mabel Muñoz, First 5 Los Angeles
Hazel Naal, LAUP
Schellee Rocher, LAUP
Rosa Valdés, LAUP

Merced
Stephanie Aguilar, First 5 Merced
Myisha Reed, First 5 Merced
Kathleem Zimmerman, First 5 Merced

Orange
Linda Nguyen, Orange County Department of Education
Limor Zimskind, Limor Consulting

San Diego
Nancy Baum, San Diego County Office of Education (SDCOE)
Gloria Corral, First 5 San Diego
Claire Crandall, SDCOE
Lynn Eldred, First 5 San Diego
Kimberly Gallo, First 5 San Diego
Lucia Garay, SDCOE

Shelby Gomez, First 5 San Diego
Steven Smith, First 5 San Diego

San Francisco
Maria Allis, First 5 San Francisco
Xavier Morales, First 5 San Francisco
Wei-min Wang, First 5 San Francisco

San Joaquin
Kristi Koumjian, Harder+Company Community Research
Kelly Mraz, First 5 San Joaquin
Lani Schiff-Ross, First 5 San Joaquin
Billi Jo Zopfi, First 5 San Joaquin

San Mateo
Jenifer Clark, First 5 San Mateo

Santa Clara
Cathy Andrade, First 5 Santa Clara
David Brody, formally with First 5 Santa Clara
Melissa Hong, First 5 Santa Clara
Jolene Smith, First 5 Santa Clara

Ventura
Dr. Heidi Christenson, Ventura County Office of Education (VCOE)
Marian Everest, VCOE Early Childhood Programs
Michell Henry, VCOE Early Childhood Programs
Carrie Murphy, VCOE Early Childhood Programs
Petra Puls, First 5 Ventura

Yolo
Margaret Bacot, City of West Sacramento
Shonna Clark, City of West Sacramento
Julie Gallelo, First 5 Yolo
Justine Jimenez, City of West Sacramento
Kristi Koumjian, Harder+Company Community Research
Ashley Mulcahy, Harder+Company Community Research
Irene Serwanga, City of West Sacramento
Talei Taganeca, City of West Sacramento

The California Department of Education's Child Development Division provided guidance in the early phases of this evaluation as F5CA attempted to align data collection efforts with RTT-ELC.

The following First 5 California staff contributed their time and expertise to this program:

Executive Staff
Diane Levin, Chief Deputy Director
Camille Maben, Executive Director

Program Management Division
Silvia Flores, Child Development Consultant
Paula Gonzales, Office Technician
Sarah Neville-Morgan, Deputy Director
Carmen Padilla, Child Development Consultant
Dr. Debra Silverman, Education Administrator
Desiree Soto, Education Administrator

Information Technology Office
Richard Flores, Senior Programmer Analyst
Riley Henderson, Staff Information Systems Analyst
Randall Ortiz, Senior Programmer Analyst
Paul Waters, System Software Specialist III Supervisor

Evaluation Office
Robert Dean, Research Program Specialist I
Dr. David Dodds, Deputy Director
Gretchen Williams, Research Program Specialist II

References

California Department of Education (CDE). 2010. *Desired Results Developmental Profile (2010) Assessment Instruments User's Guide*. Sacramento, CA: California Department of Education.

CDE. 2010a. *Desired Results Developmental Profile: Preschool*. Sacramento, CA: California Department of Education.

CDE. 2010b. *Desired Results Developmental Profile: Infant/Toddler*. Sacramento, CA: California Department of Education.

CDE 2011. *DRDP-R Technical Report*. Sacramento, CA: California Department of Education.

California Head Start Association. 2015. **"**Head Start/Early Head Start in California Data Report." Retrieved, November 19, 2015 **(**http://caheadstart.org/facts.html).

Campbell, Frances, Gabrilla Conti, James J. Heckman, Seong Hyeok Moon, Rodrigo Pinto, Elizabeth Pungello, and Yi Pan. 2014. "Early Childhood Investments Substantially Boost Adult Health." *Science* 343: 1478-1485.

Cliff, Norman. 1996. *Ordinal Methods for Behavioral Data Analysis*. Mahwah, NJ: Lawrence Erlbaum Associates.

Cohen, Jacob. 1988. *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cryer, Debby, Thelma Harms, and Cathy Riley. 2003. *All About the ECERS: A Detailed Guide in Words and Pictures to be Used with ECERS-R.* Westville, NJ: KPress Publishing.

Duncan, Greg J. and Katherine Magnuson. 2013. "Investing in Preschool Programs." *Journal of Economic Perspectives*, 27(2): 109-132.

Executive Office of the President of the United States. 2014. "The Economics of Early Childhood Investments." Washington DC, EOP.

First 5 California. 2008a. *California Power of Preschool Preliminary Statewide Evaluation Highlights FY 2006/2007*. Sacramento CA: First 5 California.

First 5 California. 2008b. *Power of Preschool Demonstration Program Progress Report Fiscal Years 2005-06 through 2007-08*. Sacramento CA: First 5 California.

First 5 California. 2009a. *Power of Preschool Program Evaluation Report*. Sacramento, CA: First 5 California.

First 5 California. 2009b. *First 5 California Power of Preschool Program Evaluation Meta-Analysis.* Sacramento, CA: First 5 California.

First 5 California. 2012a. *Request for Application for the Child Signature Program*. Sacramento, CA: First 5 California. (http://www.ccfc.ca.gov/pdf/programs/csp/rfa_1/CSP_RFA-1.pdf).

First 5 California. 2012b. *Evaluation of First 5 California's Child Signature Program*. Committee for the Protection of Human Subjects (CPHS) Protocol Number: 12-08-0632. Sacramento, CA: First 5 California

First 5 California. 2012c. *Child Signature Program Data Collection Guidebook*. Sacramento, CA: First 5 California.

First 5 California. 2014. *Evaluation of the Child Signature Program: 2012–13 School Year*. Sacramento, CA: First 5 California. (http://www.ccfc.ca.gov/pdf/research/program_reports/Evaluation_of_the_Child_Signature_Program_2012–13_School_Year.pdf)

First 5 California. 2015. *Evaluation of the Child Signature Program: 2013–14 School Year*. Sacramento, CA: First 5 California. (http://www.ccfc.ca.gov/pdf/research/program_reports/Evaluation_of_the_Child_Signature_Program_2013–14_School_Year.pdf).

Harms, Thelma, Richard M. Clifford, and Debby Cryer. 2005. *Early Childhood Environment Rating Scale: Revised Edition*. New York, NY: Teachers College Press.

Harms, Thelma, Richard M. Clifford, and Debby Cryer. 2005a. *Infant/Toddler Environment Rating Scale: Revised Edition*. New York, NY: Teachers College Press.

Harms, Thelma, Richard M. Clifford, and Debby Cryer. 2005b. *Family Child Care Environment Rating Scale: Revised Edition*. New York, NY: Teachers College Press.

Head Start. 2011. *Parent, Family, and Community Engagement Framework*. Arlington, VA: Office of Head Start, Administration for Children and Families, U.S. Department of Health and Human Services.

Heckman, James J. and Dimitriy V. Masterov, 2007. "The Productivity Argument for Investing in Young Children," *Review of Agricultural Economics,* 29(3): 446-493.

Karoly, Lynn A., Bonnie Ghosh-Dastidar, Gail L. Zellman, Michal Perlman, Lynda Ferryhough. 2007. "Prepared to Learn, The nature and Quality of Early Care and Education for Preschool-Age Children in California." Santa Monica, CA: RAND.

Kerby, Dave S. 2014. “The Simple Difference Formula: An Approach to Teaching Nonparametric Correlation.” *Innovative Teaching*, 3(1): 1-9.

McLean, M., L. Edelman, and P. Salcedo. 2011. *Users Guide to the DRDP Access for Preschool Special Education (Fourth Edition, August 2012)*. Sacramento, CA: California Department of Education, Special Education Division.

National Center for Children in Poverty. 2014. *California Demographics of Young, Poor Children.* (http://www.nccp.org/profiles/CA_profile_9.html)

Pianta, R.C., K. La Paro, and Bridget K. Hamre. 2008. *Classroom Assessment Scoring System*. Baltimore, MD: Paul H. Brookes.

Rees, Nicholas, Jingqing Chai, and David Anthony. 2012. *Right in Principle and in Practice: A Review of the Social and Economic Returns to Investing in Children*. New York, NY: United Nations Children’s Fund (UNICEF)

Romano, Jeanine, Jeffrey D. Kromrey, Jesse Coraggio, Jeff Skowronek, and Linda Devine. 2006. “Exploring Methods for Evaluating Group Differences on the NSSE and Other Surveys: Are the t-test and Cohen’s d indices the most appropriate choices?” Paper presented at the annual meeting of the Southern Association for Institutional Research, October, 14-17, 2006 Arlington, VA.

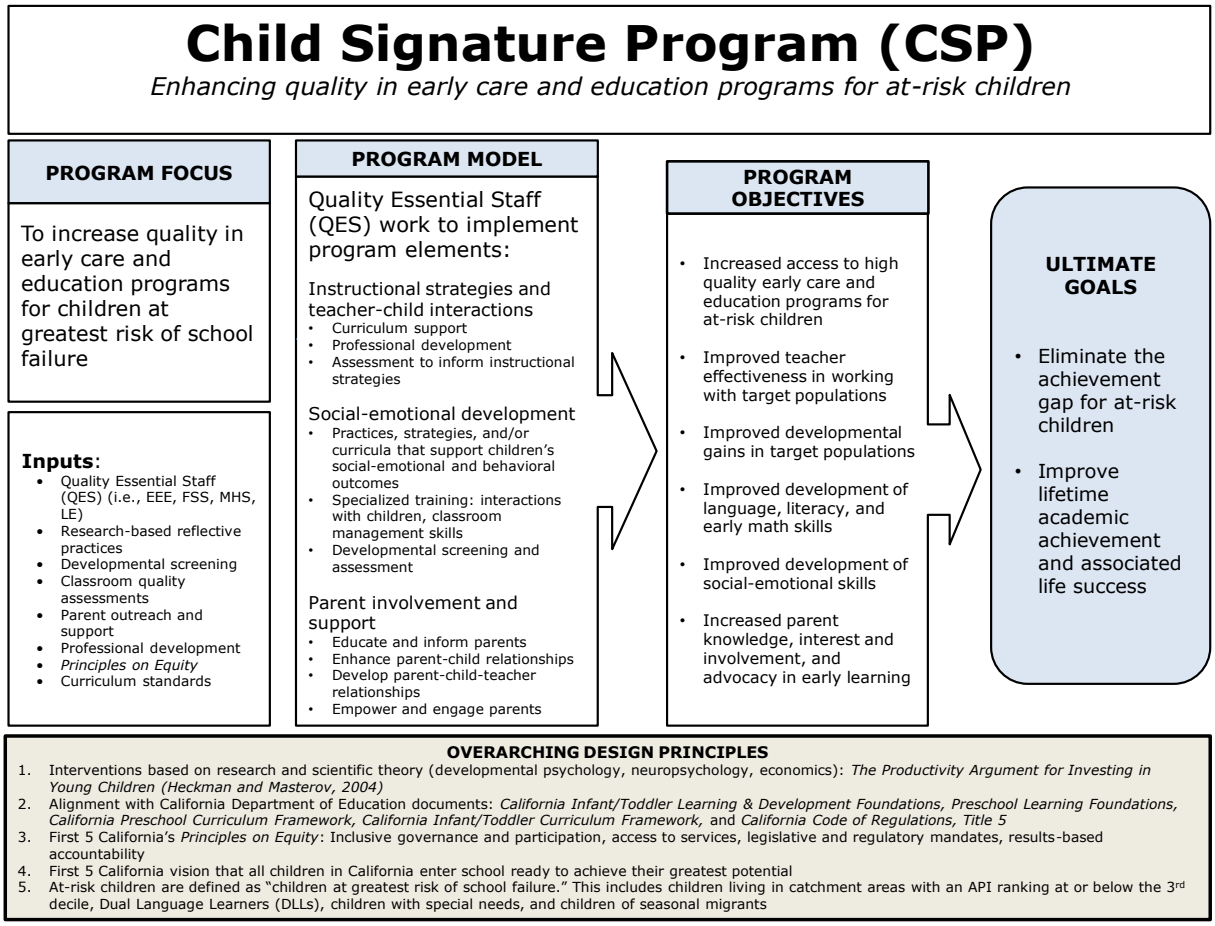
Schweinhart, Lawrence J. 2005. *The High/Scope Perry Preschool Study Through age 40: Summary, Conclusions and Frequently Asked Questions*. Ypsilanti, MI: High/Scope Press.

Schweinhart, Lawrence J. 2007. “Crime Prevention by the High/Scope Perry Preschool Program.” *Victims and Offenders*, 2:141-160.

Wuensch, Karl L. 2015. *CL: The Common Language Effect Size Statistic*. Retrieved September 17, 2015. (http://core.ecu.edu/psyc/wuenschk/docs30/CL.pdf).

Yoshikawa, Hirokazu, Christina Weiland, Jeanne Brooks-Gunn, Margaret R. Burchinal, Linda M. Espinoza, William T Gormley, Jens Ludwig, Katherine A Magnuson, Deborah Phillips, and Martha J. Zaslo. 2013. *Investing in our Future: The Evidence Base on Preschool Education*.

Appendix A: CSP Logic Model



Appendix B: Evaluation Questions Summary

Table B1. Summary of Evaluation Results to Address Outcome Questions Posed in *Attachment B* of RFA1

O.1.	Are classroom environments in CSP sites improving and meeting target quality criteria?	✓ The majority of classrooms met ERS global score and CLASS® domain score standards over the life of the program and in each program year. ✓ Average ERS global scores were above 5 (a "good" level of quality) across all school years for all age groups and classroom types. ✓ MOE preschool classrooms were able to catch up with, and then surpass, QE classrooms in terms of some ECERS subscale scores over the life of the program. ✓ Classrooms consistently met CSP quality criteria for diversity, age groups, screening, classroom ratios, group size, staff qualifications, professional development, ERS, CLASS, DRDP, and evaluation. ? It is unknown whether or not classrooms met CSP quality criteria for target groups, curriculum, articulation, health education, nutrition, nutrition education, tobacco education, physical activity, transition support, *Principles on Equity*, or budget because specific data were not collected about these criteria. Qualitative data collected through Quality Improvement Narratives (QIN) and site-level narratives suggest some of these criteria were met, but a comprehensive qualitative analysis to answer this question is beyond the scope of this report.
O.2.	Are teachers in CSP classrooms using effective teaching and classroom interaction strategies?	✓ The majority of preschool classrooms met CLASS® Pre-K standards of 5 for Emotional Support, 3 for Classroom Organization, and 2.75 for Instructional Support. These results indicate teachers, on average, are using effective classroom interaction strategies as defined by CLASS Pre-K.

O.3.	Are high-risk young children who participate in CSP demonstrating improvement in their readiness to succeed at kindergarten entry?	✓ Over the life of the program, differences in proportions of ratings in the top two developmental levels from fall to spring, across classroom quality levels, suggest children improved their readiness to succeed in kindergarten across all dimensions of development as measured by DRDP-PS.
O.4.	Is the developmental status of high risk young children who participate in CSP programs improving over time?	✓ Teacher reports suggest children experienced healthy development over the life of CSP, as indicated by consistent medium and large developmental effect sizes, regardless of classroom quality level or program year.
O.5.	Are children with special needs, Dual Language Learners (DLLs), and migrant children who attend CSP programs making developmental gains?	✓ Developmental Effect sizes for English Language Development as measured by DRDP-PS suggest DLLs were able to develop their skills with English over the life of the program and in each program year. ? This evaluation was unable to measure the overall development of DLLs, children with special needs, or children of seasonal migrants because individual level demographic data were not collected to isolate these groups for analyses.
O.6.	Are parents included in and satisfied with CSP?	✓ Data on parent satisfaction collected in 2012–13 and 2013–14 indicated very high levels of satisfaction with CSP—over 80 percent of parents were satisfied with the program. ✓ Parents of children in CSP were well informed about many aspects of their child's program. For instance, 95 percent of parents indicated they had received information about how their child was growing and developing in 2012–13. ? Overall parent participation increased 49 percent in 2013–14, but decreased from that level by 50 percent in 2014–15.

✓ Indicates results support a positive (i.e., yes) response to the research question.
? Indicates results are inconclusive

Table B2. Summary of Evaluation Results to Address Process Questions Posed in *Attachment B* of RFA1

P.1.	Are conditions that lead to and support quality early care and education increasing among programs that participate in CSP?	✓ The majority of classrooms met ERS global score and CLASS® domain score standards over the life of the program and in each program year. ✓ Average ERS global scores were 5 or above (i.e., "good" quality) across all school years for all age groups and classroom types. ✓ Classrooms consistently met CSP quality criteria for diversity, age groups, screening, classroom ratios, group size, staff qualifications, professional development, ERS, CLASS, DRDP, and evaluation.
P.2.	What strategies and services most effectively promote positive outcomes for children?	? Data on strategies were collected qualitatively through Quality Improvement Narratives (QINs) and high-level narratives on professional development and parent participation. A comprehensive development of valid categories for all types of quality improvement strategies described in CSP QINs is beyond the scope of this report.
P.3.	Are some strategies more effective for DLLs or children with special needs?	? This question could not be answered because child-level data were not collected for this evaluation. Child sub-groups could not be differentiated for analysis. Classroom-level DRDP data could not be disaggregated to assess developmental effects for different demographic groups.

P.4.	Are children with special needs being identified and receiving services as appropriate?	✓ Total counts of children identified and referred for developmental services and total counts of children receiving new developmental services for 2013–14 and 2014–15 indicate children were systematically screened for developmental delays, identified with special needs, and referred for developmental services, and that a proportion of these children also received new developmental services during the school year. ? It is not clear whether children already identified with special needs (i.e., those with and IEP or IFSP) were receiving services as intended because individual child-level demographics were not collected to define these groups for analyses. ? Child development data necessary to address this question (i.e., DRDP *access*) were not collected because local data systems were not well-integrated. QES were not able to collect necessary data because systems for sharing data did not exist or were burdened with regulations or practices that complicated collecting, storing, and sharing information about protected groups (i.e., children with special needs).
P.5.	What are the most effective outreach strategies for parents?	? Although quantitative data were collected about parent participation, only qualitative data were collected through Quality Improvement Narratives (QIN) and site-level narratives on outreach, engagement, and support activities provided to parents. An adequate development of categories for all the different possible types of outreach, engagement, and support activities described in high-level qualitative narratives is beyond the scope of this report.

✓ Indicates results support a positive (i.e., yes) response to the research question.
? Indicates results are inconclusive

Appendix C: Evaluation Challenges and Lessons Learned

Evaluation of CSP required flexibility for the tasks of evaluation design, data collection, and reporting. Following are some of the key areas where original evaluation plans were modified during program implementation and how these modifications may have impacted final evaluation results.

Evaluation Design

F5CA chose to evaluate CSP using a quasi-experimental research design for two reasons. First, the ultimate goal of eliminating the achievement gap for at-risk children implies comparison of educational outcomes between more and less privileged groups (Appendix A, CSP Logic Model). Because CSP was designed to serve low-income families and children, a comparison group (i.e., control group) of families and children not participating in the program was unavailable as part of an experimental design. Second, because CSP adopted two quality levels for its classrooms, QE and MOE, it was possible to think of these two types of classrooms as part of a natural, quasi-experiment allowing comparison of inputs and outcomes for the two quality levels. However, during real-world program implementation, challenges in program targeting and the role of QES in serving both QE and MOE classrooms blurred some of the contrasts intended to be captured by the quasi-experimental evaluation design. The difference between treatment (QE) and control (MOE) groups was most affected by program requirements difficult to implement by CSP counties with regard to program targeting and the role of QES.

Program Targeting

The original intent of CSP was that all classrooms would be QE. One program requirement for QE classrooms was location in lower-performing areas in public school catchment areas with scores in the bottom three deciles of the API—only about half of the classrooms in the program met this criterion. However, counties also were required to serve 90 percent of children in areas targeted by the predecessor program, PoP. Typically, counties could not easily meet both requirements because they did not have enough PoP classrooms in lower-performing areas as evidenced by the API. Additionally, PoP and prior F5CA preschool initiatives may have raised API scores in previous low-decile areas.

Quality Essential Staff

Another issue for counties was the costs and logistics of hiring QES to work with all participating classrooms. Shifting supports and funding to align with original requirements of CSP might have resulted in eliminating services for prior PoP classrooms—working against basic goals of both PoP and CSP. F5CA responded by relaxing initial program requirements and including the MOE quality level so counties could direct supports and funding to areas with more need. The revised program

requirements allowed counties to serve more children. As a result, the primary factor differentiating treatment and control groups was the absence or presence of QES. Qualitative analyses of teacher narratives from both QE and MOE classrooms over all three years revealed three important facts: 1) QES did not work strictly in QE classrooms; 2) over the life of the program, QES tended to work more with the MOE classrooms; and 3) QES served MOE classrooms indirectly as they developed systems and processes to support sites (e.g., professional development for all teachers at a site). In light of the real-world need for QES services in QE and MOE classrooms, this complexity made it difficult to detect differences between QE and MOE classrooms.

Data Collection

Evaluation data collection involved multiple challenges, including the need to maintain flexibility for the frequency and volume of assessments performed, linking evaluation questions to data that were feasible to collect, achieving sampling size sufficient to detect statistical patterns in assessment data, and coordinating data management across multiple local data systems and the CSP statewide data system.

Flexibility for Assessments

For CSP, flexibility for assessment data collection was important to achieve the program evaluation. Particular examples include:

- **Revisions to CLASS® and ERS Observation Schedules:** In the first evaluation plan, all CSP classrooms were to receive pre- and post-observations with CLASS and ERS instruments each year. Based on feedback from counties about the cost and complex logistics for the first plan, F5CA revised requirements for CLASS and ERS observations. As a result, the original annual pre-post quasi-experimental design was adapted so that evaluation classrooms collected CLASS and ERS observation data once per year.

- **Reliable Raters:** During the first year of the program, county evaluators were unclear about what constituted a “reliable outside rater” and this may have affected ERS scores for 2012–13. To address this concern (and others), F5CA developed the *CSP Data Collection Guidebook* to align state and local research activity, processes, and systems.

- **DRDP *access*:** From a local perspective, LEs found it difficult to collect and report DRDP *access* results because there were no clear data sharing practices between CSP staff and primary special education service providers who typically collect data on children with special needs. A handful of CSP counties were able to link developmental screening and assessment activity into coherent systems to support the sharing of DRDP *access* data during CSP.

Evaluation Question Specificity and Data Collection Feasibility

Addressing all evaluation questions sufficiently with data collected was challenging for several reasons: 1) some data collected did not specifically address questions posed in the *Request for Application for the Child Signature Program*, *Attachment B* as designed prior to program launch; 2) some evaluation questions contained multiple nested questions with an overbroad focus; 3) evaluation questions and units of analysis were sometimes mismatched; and 4) some data collection instruments may not have been sensitive enough to detect significant differences in quality between CSP classrooms.

As an example of mismatch between evaluation questions, feasibility of data collection, and units of analysis, *Attachment B*, Question O.5. asks "[a]re children with special needs, DLLs, and migrant children who attend CSP programs making developmental gains?" This question is best answered with individual child-level data and demographics. However, child-level data were not collected during CSP because of human subject protection requirements for child-level data, information technology capacity for building the CSP data system, and the burden of data collection for county evaluators.

Sampling and Assessment Data

During planning for data collection, statistical power analyses indicated an annual sample size between $n = 200$ and $n = 250$ classrooms for detecting small to medium effect sizes (Cohen's d, 0.2 to 0.5). However, with the exception of Instructional Support Domain scores in 2014–15, CLASS® scores did not show statistically significant differences between QE and MOE classrooms. There are three possible explanations: 1) QE and MOE classrooms were not different; 2) sample sizes were too small to detect existing differences; or 3) classrooms were different, but CLASS scoring methodology was not sensitive enough to enumerate differences in quality. By contrast, data collected with ERS instruments were able to detect significant differences in structural quality, and data from DRDP assessments were sensitive enough to detect significant differences in developmental effect sizes.

Data Systems

Data collection for the evaluation of CSP was challenging because of different data systems. CSP counties collected data using a constellation of locally run data systems. Data collection, entry, reporting, and storage processes were different across CSP counties. Thus, it was difficult to design a data system to support statewide evaluation while meeting needs of CSP counties. During the first year of CSP, F5CA's Information Technology Office was able to modify the CSP Data Profile System, used for baseline data in preparation of the program's launch, to collect evaluation data outlined in the *CSP Data Collection Guidebook*. Though the CSP data system served as a structured repository of uniform data, participating counties had to perform manual entry of data already maintained in local data systems. Most counties hired additional staff to transfer data from local systems to the state system. Some counties were able to create scripts

to reduce data entry labor while transferring data into the statewide CSP data system. Because of the relatively short life of the program, three years, it was difficult to develop a more streamlined and labor-efficient method of sharing data between local and state data systems.

Appendix D: Supplemental Data Tables and Figures for Report and Analyses of Classroom Developmental Effect Sizes

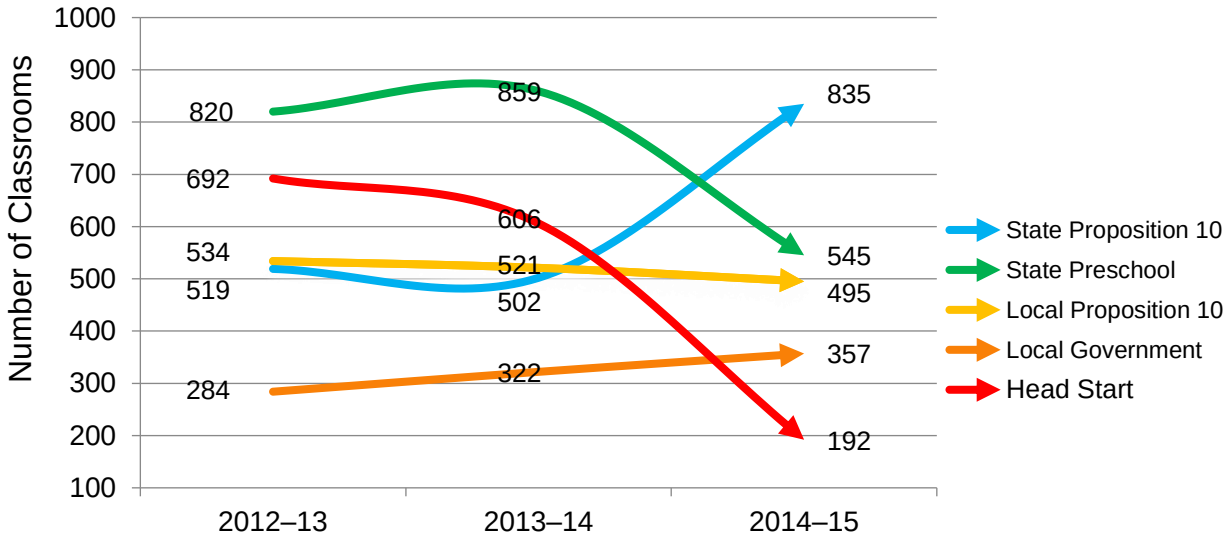
Funding Sources

Table D1. Classrooms by Funding Source: 2014–15

Funding Source	QE Classrooms Funded (*N* = 137)	Percent	MOE Classrooms Funded (*N* = 1,213)	Percent	Total Classrooms Funded (*N* = 1,350)	Percent
State Proposition 10	90	66%	745	61%	835	62%
State Preschool (i.e., CSPP)	69	50%	476	39%	545	40%
Local Proposition 10	68	50%	427	35%	495	37%
Local Government	19	14%	338	28%	357	26%
Other	19	14%	293	24%	312	23%
Federal Other	23	17%	262	22%	285	21%
Head Start	19	14%	173	14%	192	14%
State General Childcare	20	15%	119	10%	139	10%
State Alternative Payment	8	6%	117	10%	125	9%
External Gifts or Donations	3	2%	11	1%	14	1%
Local Other	6	4%	9	1%	15	1%
Early Head Start	4	3%	2	<1%	6	<1%
External/Non-Profit Organization	4	3%	1	<1%	5	<1%
External Foundation	2	1%	4	<1%	6	<1%
State Other	0	0%	5	<1%	5	<1%

Note: Classrooms may have more than one funding source.

Figure D1. Change in Classrooms Funded by Top 5 Funding Sources 2012–2015

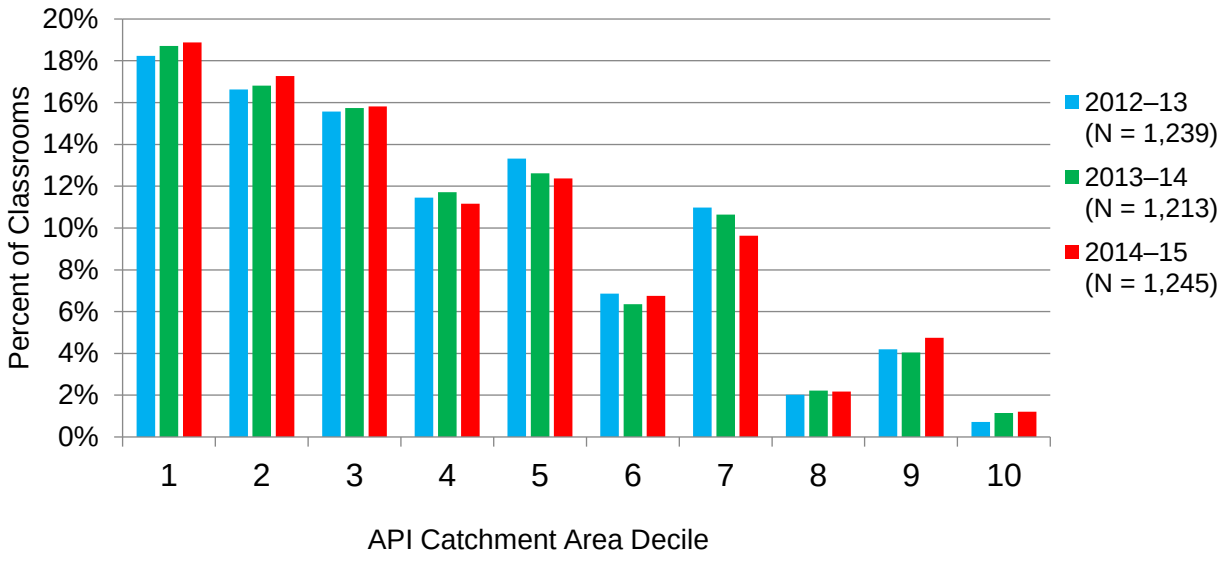


Location of Classrooms by API Catchment Area

Table D2. Classrooms by API Decile and Classroom Quality Level: 2014–15

API Decile	QE Classrooms (*N* = 128)		MOE Classrooms (*N* = 1,117)		All classrooms (*N* = 1,245)	
1	27	21%	208	19%	235	19%
2	34	27%	181	16%	215	17%
3	24	19%	173	15%	197	16%
4	19	15%	120	11%	139	11%
5	12	9%	142	13%	154	12%
6	4	3%	80	7%	84	7%
7	0	0%	120	11%	120	10%
8	2	2%	25	2%	27	2%
9	3	2%	56	5%	59	5%
10	3	2%	12	1%	15	1%

Figure D2. Classrooms by API Deciles and Classroom Quality Level: Three School Years, 2012–15



Special Populations

Table D3. Special Populations Served by Classroom Quality Level: 2014–15

	Children who are DLL			Children with Special Needs			Children of Seasonal Migrant Families		
	Number	Percent of Total Children Within Quality Level	Percent of Total Target Group	Number	Percent of Total Children Within Quality Level	Percent of Total Target Group	Number	Percent of Total Children Within Quality Level	Percent of Group
QE	1,525	63%	11%	97	4%	10%	15	1%	48%
MOE	12,213	57%	89%	834	4%	90%	16	<1%	52%
All	13,738	58%	100%	931	4%	100%	31	<1%	100%

Note: N = 23,675 total children served, N = 21,245 children served through MOE, and N = 2,430 children served through QE classrooms.

QE classrooms served 11 percent of children who are DLLs and 10 percent of children with special needs; MOE classrooms served 89 percent and 90 percent respectively. QE classrooms served higher proportions of DLLs (63 percent compared to 57 percent) but the same proportion (4 percent) of children with special needs in 2014–15.

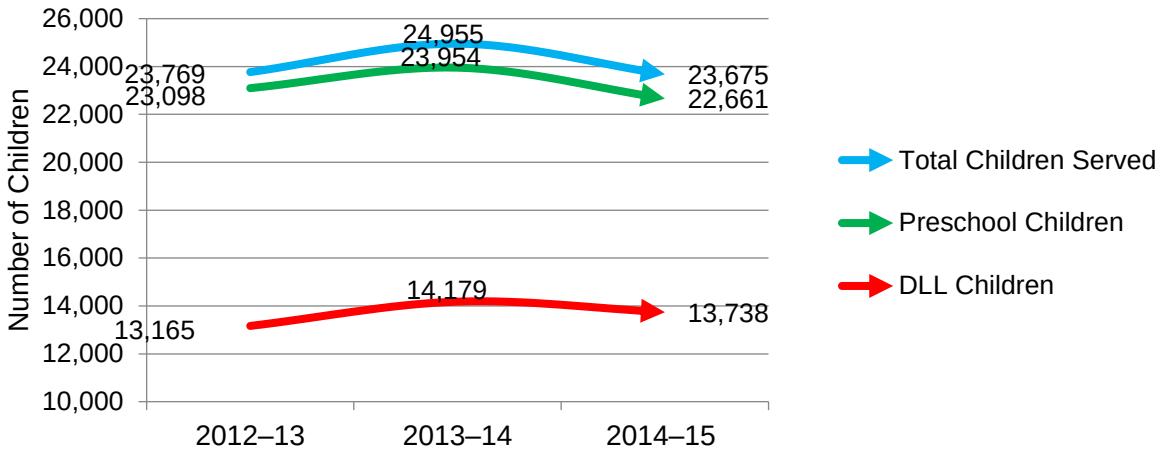
Table D4. Primary Language of DLLs Served: 2014–15

Language	Number	Percent of DLL (N = 12,492[a])	Percent of all Children Served (N = 23,675)
Spanish	9,550	70%	40%
Chinese	1,362	11%	6%
Other	96	1%	<1%
Filipino	120	1%	1%
Vietnamese	160	1%	1%
Korean	103	1%	<1%
Arabic	102	1%	<1%
Russian	66	1%	<1%
Total DLL	13,738[a]	100%	58%

a. Percents are for N = 12,492 children reported by language among N = 13,738 total DLL served.

Special Populations: Changes in Counts Across Years

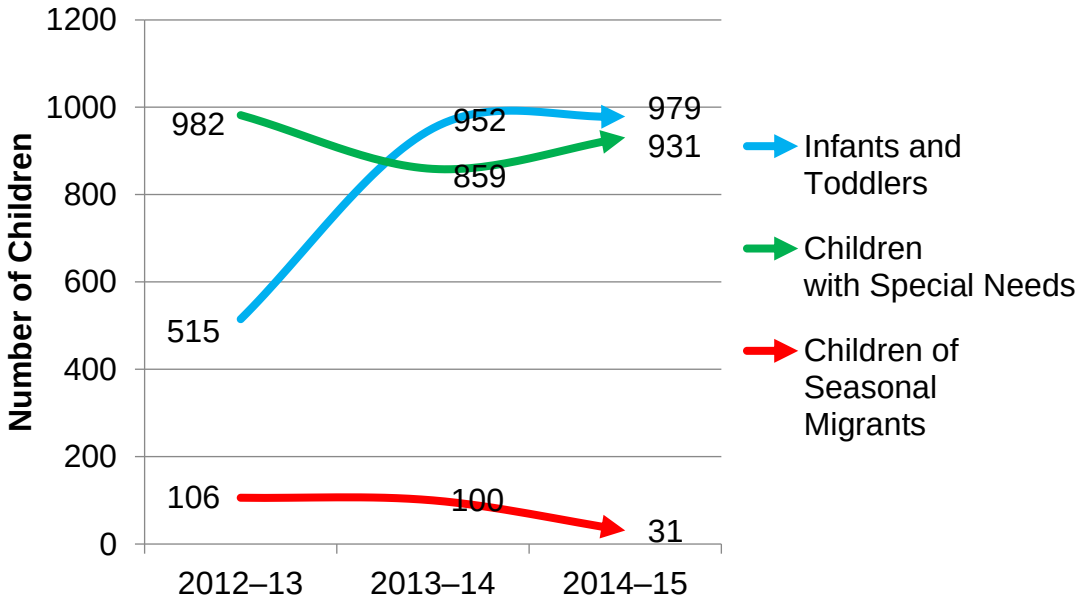
Figure D3. Change in Children Served by Target Group Across School Years for Target Groups of 10,000 or More



Note: Graph includes only target groups of 10,000 children or more

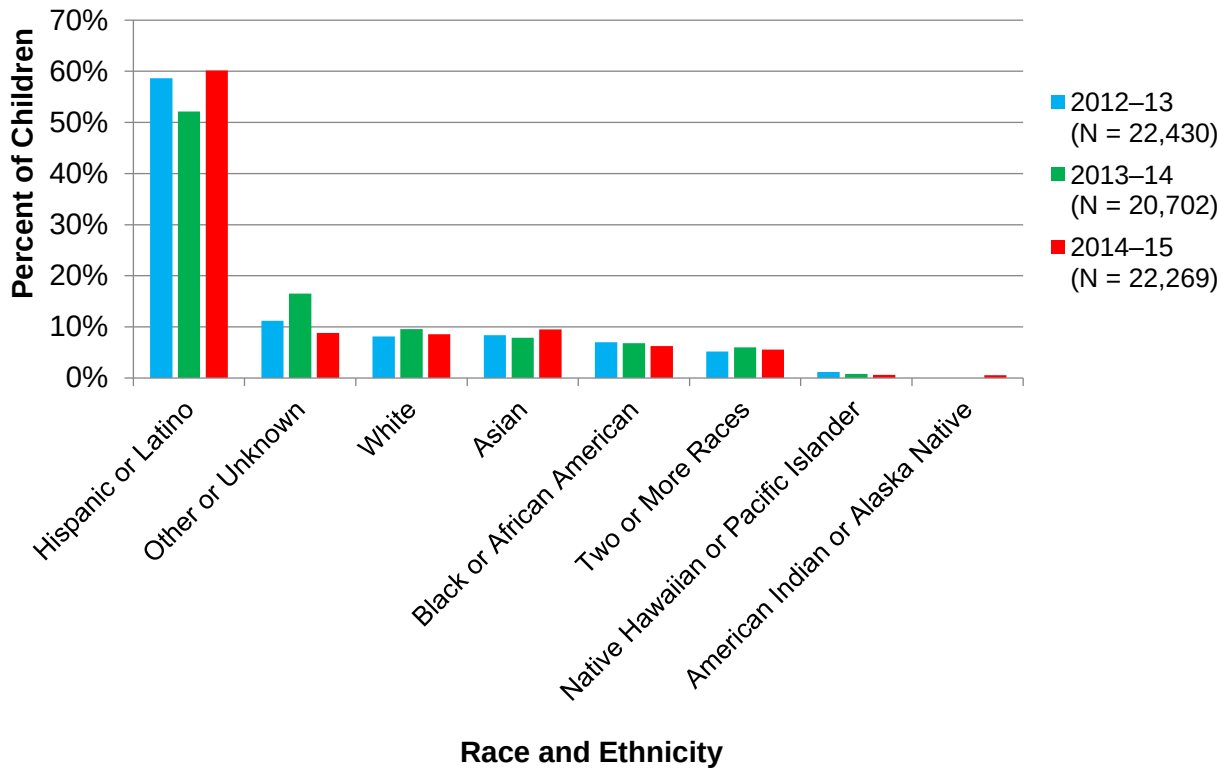
CSP served fewer total children in 2014–15 than in previous school years, and served more DLLs than in 2012–13, but fewer DLLs than in 2013–14.

Figure D4. Change in Children Served by Target Group Across School Years for Target Groups of 1,000 or Fewer



Note: Graph includes only target groups of 1,000 children or fewer.

Figure D5. Children Served by Race and Ethnicity: Three School Years (2012–2015)



Structural Quality Indicators: Ratio and Group Size

Table D5a. Teacher to Student Ratios by Classroom Quality Level: 2014–15[8]

	Preschoolers		Toddlers		Infants	
	Mean Ratio	*N*	Mean Ratio	*N*	Mean Ratio	*N*
QE	1:8	103	1:4	12	1:3	8
MOE	1:8	984	1:3	20	1:3	9
All	1:8	1,087	1:3	32	1:3	17

[8] A more complete analysis of ratio data could involve calculation of the percent of classrooms meeting teacher or provider-child ratio and classroom group size quality criteria. However, these data are difficult to categorize for analysis because of the complexity of interacting quality criteria and program standards. Licensing, location of the classroom, local policy, funding sources, education and qualifications of teaching staff, program type, etc., all influence the ratio and group size standards CSP classrooms must meet. Mean ratio and group sizes, on the other hand, are useful because they show how much CSP classrooms tend to meet the range of standards.

Table D5b. Provider to Student Ratios by Classroom Quality Level: 2014–15

	Preschoolers		Toddlers		Infants	
	Mean Ratio	*N*	Mean Ratio	*N*	Mean Ratio	*N*
QE	1:6	26	1:3	4	1:4	3
MOE	1:7	564	1:3	18	1:3	9
All	1:7	590	1:3	22	1:3	12

Note: Mean ratios are rounded to the nearest whole number.

Table D6. Mean Classroom Group Sizes by Classroom Quality Level and Age Groups: 2014–15

	Preschoolers		Toddlers		Infants		Total Children Served	
	Mean Group Size	*N*	Mean Group Size	*N*	Mean Group Size	*N*	Mean Group Size	*N*
QE	20	110	9	16	8	10	19	126
MOE	19	1,054	6	101	5	36	20	1,079
All	19	1,164	6	117	5	46	20	1,205

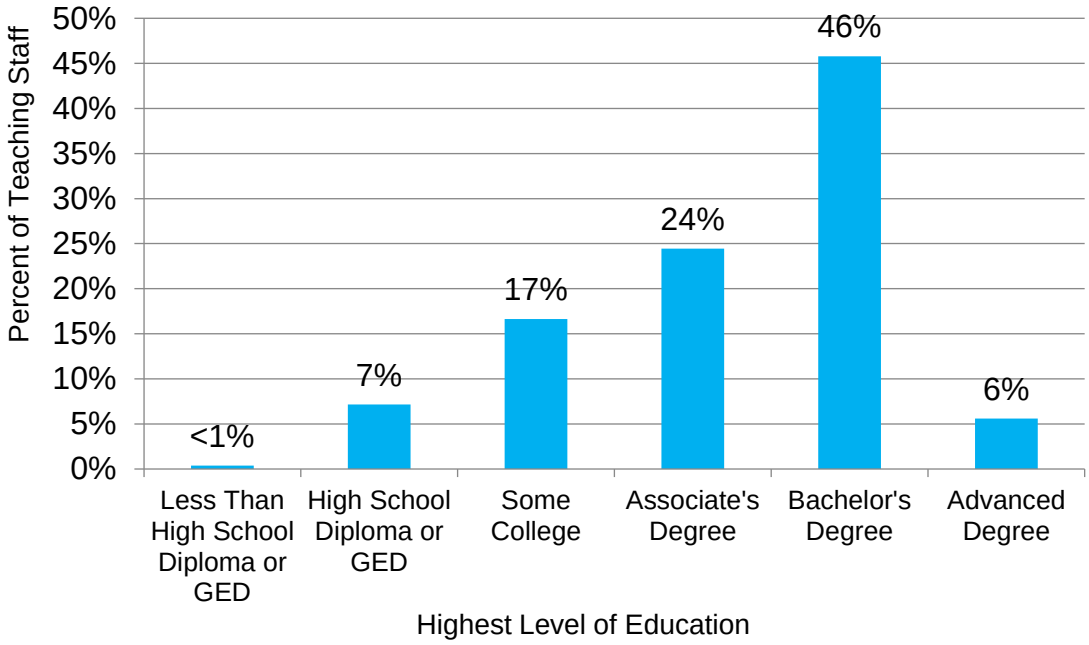
Structural Quality Indicators: Teacher/Provider Qualifications

Table D7. Teaching Staff by Highest Level of Education and Classroom Quality Level: 2014–15

	QE		MOE		All Classrooms	
	Number	Percent	Number	Percent	Number	Percent
Less Than High School Diploma or GED	1	<1%	6	<1%	7	<1%
High School Diploma or GED	5	2%	127	8%	132	7%
Some College	27	12%	280	17%	307	17%
Associate's Degree	60	27%	391	24%	451	24%
Bachelor's Degree	117	53%	728	45%	845	46%
Advanced Degree	10	5%	93	6%	103	6%
Total	220	12%	1,625	88%	1,845	100%

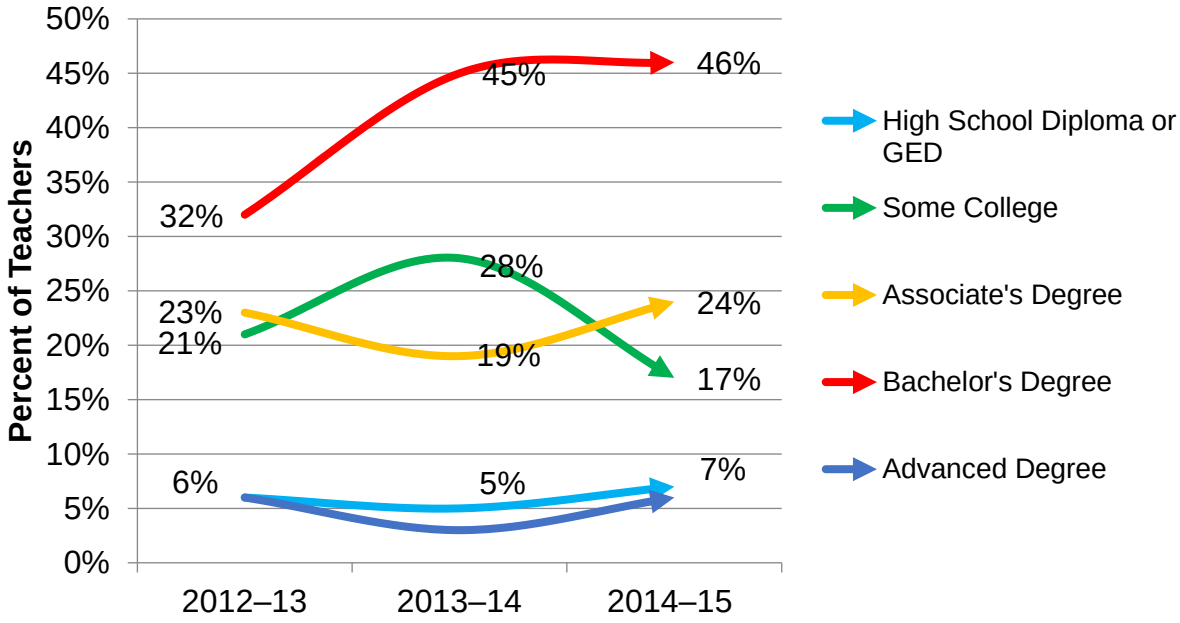
Note: CSP teaching staff can work in multiple classrooms. Data used to create this table were collected as classroom-level data. Percents are based on N = 1,845 teaching staff records with data on highest level of education for approximate N = 1,162 teaching staff.

Figure D6. Teaching Staff by Highest Level of Education: 2014–15



Note: Percents are for N = 1,845 records with data on highest level of education.

Figure D7. Change in Percent of Teaching Staff by Highest Level of Education Across School Years



Note: Percents are for N = 1,845 records with data on highest

Table D8. ECE or CD Units and Degrees by Classroom Quality Level: 2014–15

	QE	MOE	All Classrooms
Mean ECE or CD Units Per Classroom[a]	71.30	53.91	55.89
Number of ECE or CD Degrees	140	652	792
Estimated Percent ECE or CD Associate's Degrees[b]	39%	47%	45%
Estimated Percent ECE or CD Bachelor's Degrees	55%	45%	47%
Estimated Percent Advanced ECE or CD Degrees	6%	8%	7%
Estimated Percent of Teaching Staff with ECE or CD Degrees[c]	59%	36%	38%

a. N = 1,071 classrooms (MOE = 949, QE = 122) with data on ECE units held by teaching staff. Difference in mean ECE or CD units between QE and MOE is statistically significant at the p<.0001 level.
b. N = 2,067 teaching staff records (MOE = 1,829, QE = 238). Teachers may be duplicated across classroom quality levels.
c. Percents based on N = 2,067 (MOE = 1,829, QE = 238) teaching staff records for approximate N = 1,162 teaching staff working across CSP classroom quality levels.

Figure D8. Change in ECE or CD Units Held by Teaching Staff Across School Years

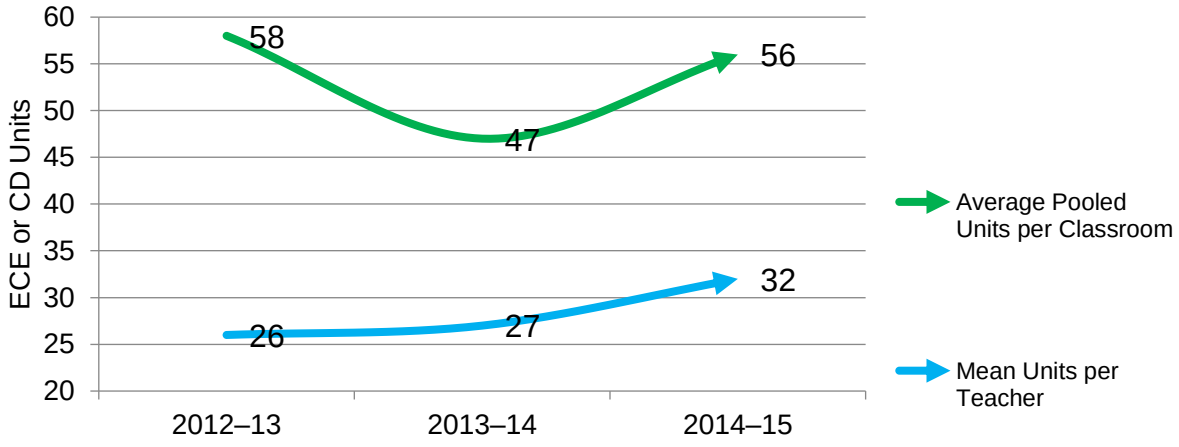
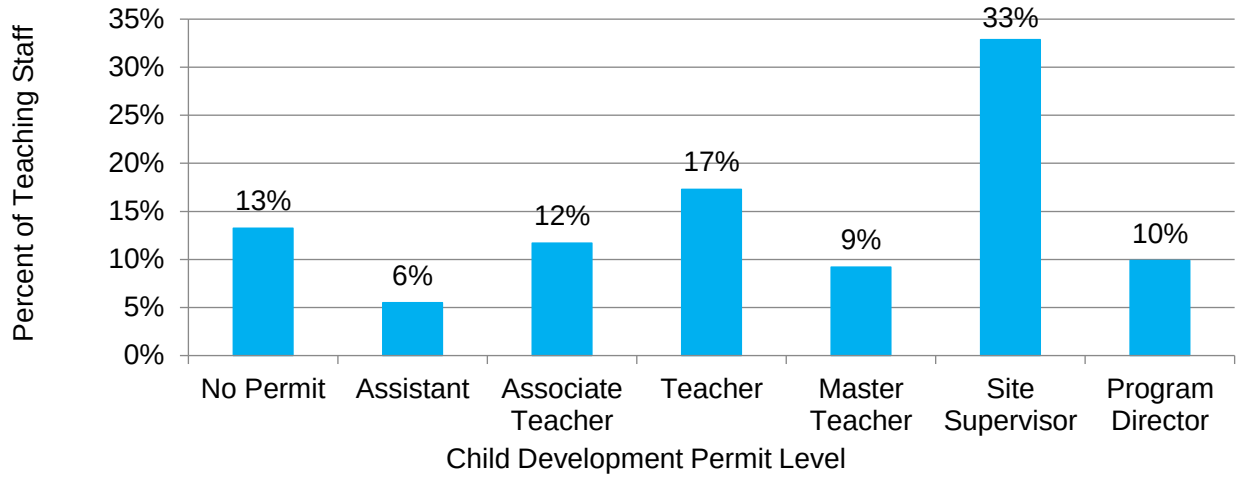


Figure D9: Teaching Staff by Child Development Permit Level: 2014–15



Note: Percents are based on approximate N = 1,162 teaching staff.

Figure D10. Participation in Professional Development Across School Years

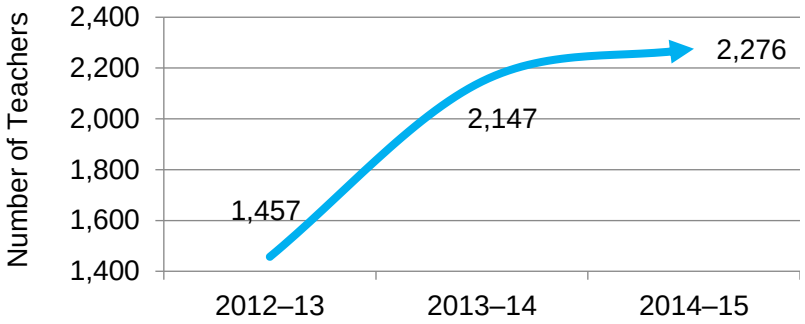
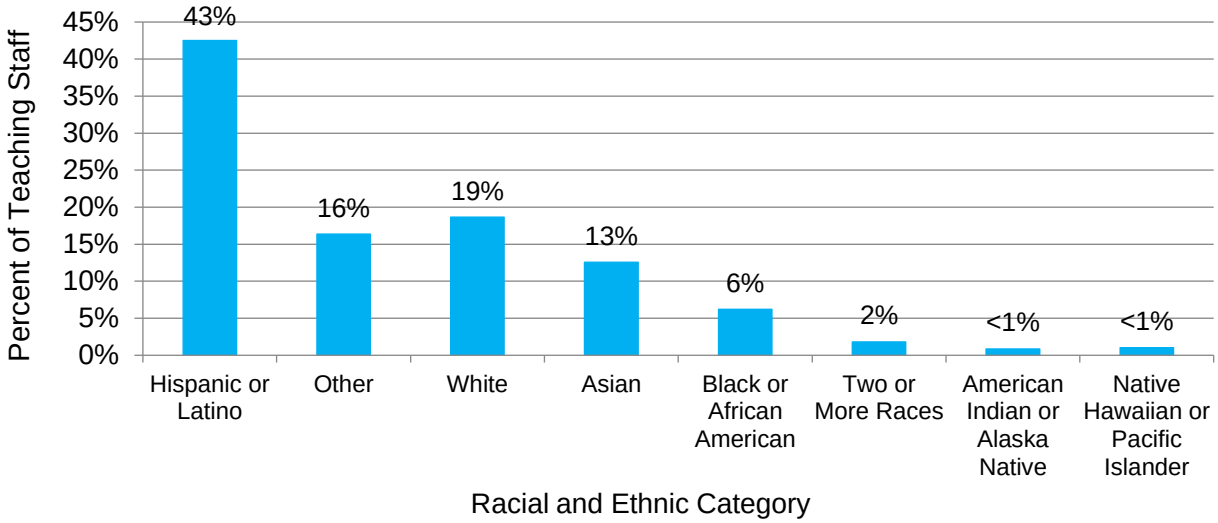
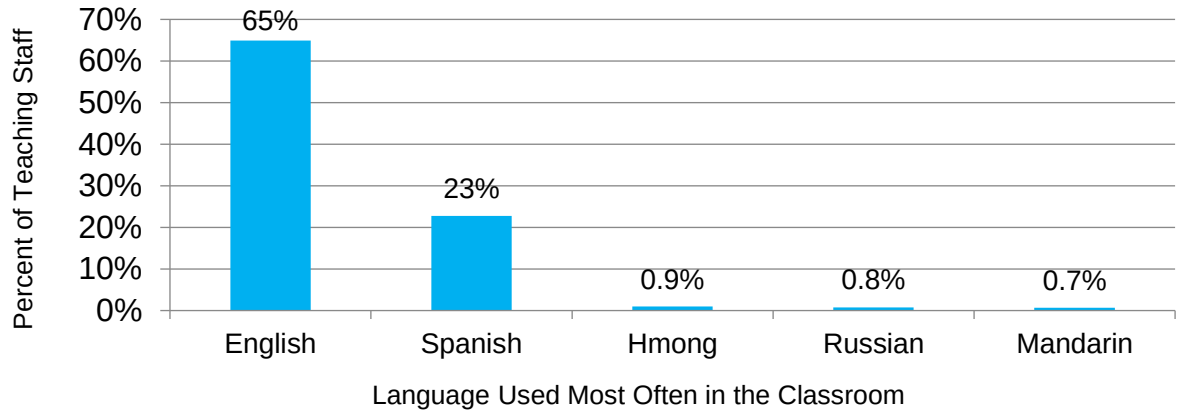


Figure D11. Classroom Teaching Staff by Racial and Ethnic Category: 2014–15



Note: Percents are based on an approximate N = 1,162 teaching staff.

Figure D12. Teaching Staff by Language Used Most Often in the Classroom: 2014–15



Note: Percents are based on an approximate N = 1,162 teaching staff.

Classroom Quality

Environment Rating Scales

Table D9. Comparison of ERS Global Scores by Classroom Quality Level: 2014–15

	QE				MOE				Mann-Whitney *U* test	
	Median	Min	Max	*N*	Median	Min	Max	*N*	*p*-Value	CL_{QE}
ECERS	5.6	3.6	6.7	111	5.6	2.5	6.7	414	0.2305	48%
ITERS	5.6	2.3	6.9	18	5.7	2.6	6.6	15	0.4496	51%
FCCERS	5.1	4.3	5.9	4	5.8	4.2	6.8	22	0.2950	41%

Note: CL_{QE} is the probability a randomly selected QE classroom will have a higher ERS global score than a randomly selected MOE classroom. Data are for all classrooms receiving and reporting ERS observations during the 2014–15 school year.

In Table D9, nonparametric Mann-Whitney *U* tests[9] did not detect statistically significant differences in the distribution of ranked global scores between QE and MOE classrooms assessed with any ERS tool. Common language effect size (CL)[10] for these tests

[9] Nonparametric tests (i.e., Mann-Whitney *U*) are more appropriate to evaluate these results because the data violate the following assumptions for parametric hypothesis testing (i.e., student's *t*-test): 1) data are skewed and not normal, 2) results for ERS and CLASS appear to be continuous but are actually derived from ordinal measures, and 3) where sampling is a factor, only MOE classrooms were randomly selected into the sample population and classrooms were not randomly selected into treatment and control groups.

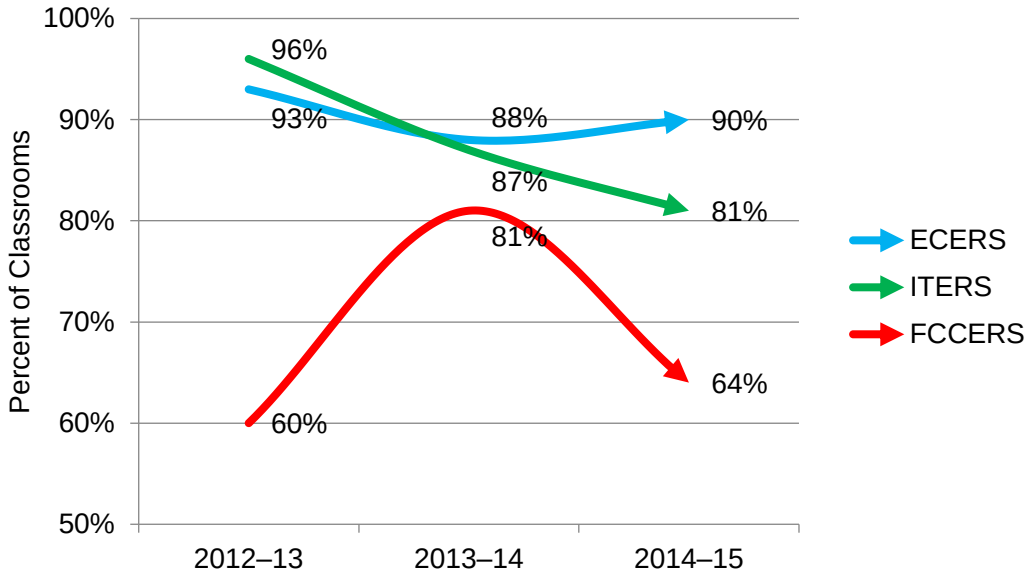
[10] The common language effect size (CL) indicates the probability a randomly selected classroom from one group received a higher score on some measure than a randomly selected classroom from another group. CL is calculated by dividing the obtained Mann-Whitney *U* statistic by the product of the sample sizes of the two groups being compared. CL is a more appropriate effect size measure for nonparametric statistical tests than Cohen's *d* (Kerbey 2014); however, CL effect sizes of 56 or 44 percent, 64 or 36 percent, and 72 or 28 percent, correspond to Cohen's *d* effect sizes of 0.2 (small), 0.5 (medium), and 0.8 (large) (Wuensch 2015). CL_{QE} is calculated using this same logic but expresses the probability a randomly selected QE classroom will have a higher score than a randomly selected MOE classroom.

revealed MOE classrooms outperformed QE classrooms for FCCERS, but distributions of ranked scores for ECERS and ITERS were not different.

Table D10. Distribution of Classrooms Meeting ERS Global Score Standards: Three School Years (2012–2015)

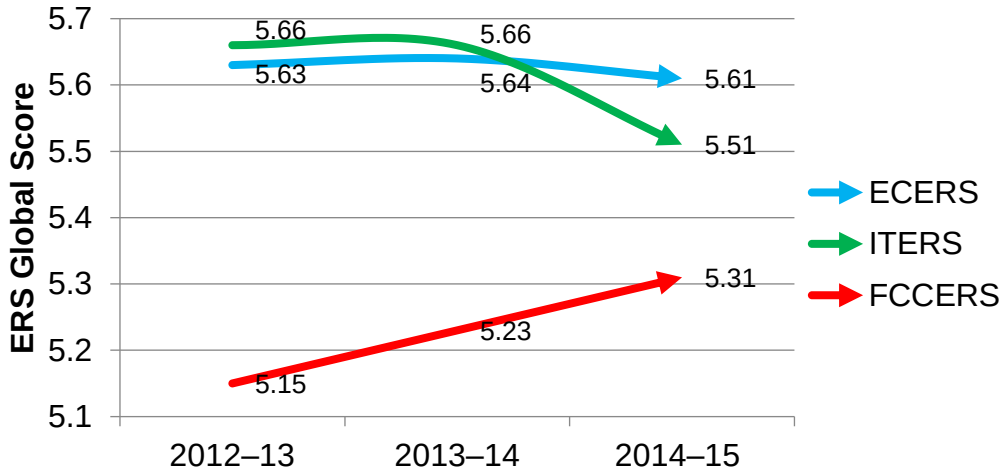
	QE		MOE		All Classrooms	
	≥5	N	≥5	N	≥5	N
ECERS	83%	230	91%	1,548	90%	1,778
ITERS	79%	33	93%	41	86%	74
FCCERS	69%	9	78%	98	77%	107

Figure D13. Classrooms Meeting ERS Global Score Standards Across School Years: 2012–2015



Note: Results for ITERS and FCCERS should be interpreted with caution because small group sizes across school years could have contributed to the variability in scores.

Figure D14. Change in Average ERS Global Scores Across School Years



Note: Average ERS global scores for 2012–13 include baseline scores collected prior to 2012.

Table D11 shows the ECERS subscale differences between QE and MOE classrooms. In terms of CL_{QE} effect size, MOE classrooms generally showed higher levels of quality in 2014–15 across subscales with the exception of the Interaction subscale. For Interaction, CL_{QE} indicates a 63 percent probability a QE classroom would have a higher score than an MOE classroom. For Activities, CL_{QE} indicates a 77 percent probability an MOE classroom would have a higher score than a QE classroom—a large effect. For Space and Furnishings, CL_{QE} indicates a 70 percent probability an MOE classroom would have a higher score than a QE classroom. For Parents and Staff, CL_{QE} indicates a 64 percent probability an MOE classroom would have the higher score.

Table D11. Comparison of ECERS Subscale Scores by Quality Level: 2014–15[11]

	QE (N = 41)			MOE (N = 35)			Mann-Whitney *U* test	
	Median	Min	Max	Median	Min	Max	*p*-Value	CL_{QE}
Space and Furnishings	5.0	3.0	6.9	5.9	3.5	7.0	0.0018**	30%
Personal Care Routines	3.3	1.8	5.8	3.3	1.8	5.8	0.4833	50%
Language Reasoning	5.8	3.0	7.0	6.0	3.8	7.0	0.0610	44%
Activities	5.4	2.8	6.9	6.0	5.1	7.0	<0.0001***	23%
Interaction	6.8	4.6	7.0	6.4	5.2	7.0	0.0269*	63%
Program Structure	5.3	3.0	7.0	5.5	4.0	6.8	0.3126	47%
Parents and Staff	6.5	1.3	6.8	6.7	5.3	7.0	0.0167*	36%

Note: Data are for evaluation classrooms only.
* $p<.05$, ** $p<.01$, *** $p<.001$

Table D12 summarizes results of statistical tests for differences in ECERS subscale scores across all three years of CSP. Six tests, 4 in 2012–13, 1 in 2013–14, and

[11] These analyses were not performed for ITERS or FCCERS results due to low sample sizes (< 5) for the 2014–15 school year. Analyses for these instruments using the full 3-year dataset appear below.

another in 2014–15, supported the evaluation hypothesis in that QE classrooms showed higher levels of quality than did MOE classrooms. However, 6 additional tests, 2 in 2012–13, and 4 in 2014–15, did not support the hypothesis.

Table D12. Summary of Findings Across Program Years for Analyses of Differences in Classroom Quality by ECERS Subscales

	2012–13	2013–14	2014–15	2012–15 Pooled
Space and Furnishings	QE > MOE	No Difference	QE < MOE	QE < MOE
Personal Care Routines	QE > MOE	No Difference	No Difference	No Difference
Language Reasoning	No Difference	No Difference	QE < MOE	No Difference
Activities	QE < MOE	No Difference	QE < MOE	QE < MOE
Interaction	QE > MOE	No difference	QE > MOE	No Difference
Program Structure	QE > MOE	QE > MOE	No Difference	QE < MOE
Parents and Staff	QE < MOE	No Difference	QE < MOE	No Difference

Note: Notable differences between quality levels are associated with Cohen's *d* effect sizes of .2 or larger.

Table D13 shows Mann-Whitney *U* tests detected statistically significant differences in ranked ECERS subscale scores for Space and Furnishings, Activities, and Program Structure subscales with small CL_{QE} effect sizes.

Table D13. Comparison of ECERS Subscale Scores by Quality Level: Three School Years (2012–2015)

	QE (N = 216)			MOE (N = 257)			Mann-Whitney *U* test	
	Median	Min	Max	Median	Min	Max	*p*-Value	CL_{QE}
Space and Furnishings	5.1	2.8	7.0	5.4	2.9	7.0	0.0021**	42%
Personal Care Routines	3.2	1.3	6.2	3.0	1.3	5.8	0.1083	53%
Language Reasoning	5.7	2.3	7.0	5.7	3.0	7.0	0.4137	51%
Activities	5.6	2.6	7.0	5.8	2.8	7.0	0.0030**	43%
Interaction	6.4	2.0	7.0	6.4	2.8	7.0	0.2914	51%
Program Structure	5.3	1.3	7.0	5.3	2.0	7.0	0.0074**	44%
Parents and Staff	6.3	1.3	7.0	6.3	1.5	7.0	0.2154	48%

Note: Data are for all evaluation classrooms receiving ECERS observations across all school years.
* $p<.05$, ** $p<.01$, *** $p<.001$

In Table D14, Mann-Whitney *U* tests detected one statistically significant difference in ranked subscale scores for Language Reasoning. CL_{QE} for this test indicates a 67 percent probability an MOE classroom would have a higher Language Reasoning score than a QE classroom.

Table D14. Comparison of ITERS Subscale Scores by Quality Level: Three School Years (2012–2015)

	QE (N = 31)			MOE (N = 15)			Mann-Whitney *U* test	
	Median	Min	Max	Median	Min	Max	*p*-Value	CL_{QE}
Space and Furnishings	5.6	1.8	7.0	5.6	3.2	6.4	0.4439	49%
Personal Care Routines	4.0	1.2	6.5	4.0	2.3	4.8	0.3623	53%
Language Reasoning	5.3	2.0	7.0	5.3	4.3	7.0	0.0316*	33%
Activities	4.8	0.2	6.2	4.7	4.1	5.7	0.4953	50%
Interaction	6.0	0.5	7.0	6.5	3.3	7.0	0.2368	43%
Program Structure	4.5	1.0	7.0	5.3	2.8	6.3	0.1144	39%
Parents and Staff	6.4	3.7	7.0	6.6	5.3	7.0	0.0767	37%

Note: Data are for all infant/toddler classrooms receiving ITERS observations across all school years.
* *p*<.05, ***p*<.01, ****p*<.001

Table D15 compares distributions of ranked FCCERS subscale scores across classroom quality levels for all FCC homes reporting subscale scores over the life of CSP. Mann-Whitney *U* tests detected one statistically significant difference for the Parents and Staff subscale. CL_{QE} for this result indicates a 72 percent probability a QE classroom would have a higher score than an MOE classroom.

Table D15 Comparison of FCCERS Subscale Scores by Quality Level: Three School Years (2012–2015)

	QE (N = 31)			MOE (N = 15)			Mann-Whitney *U* test	
	Median	Min	Max	Median	Min	Max	*p*-Value	CL_{QE}
Space and Furnishings	5.0	1.8	5.2	5.5	3.0	6.8	0.0686	32%
Personal Care Routines	3.0	1.7	4.0	3.3	1.2	7.0	0.2199	41%
Language Reasoning	6.3	2.0	7.0	5.7	2.3	7.0	0.0753	67%
Activities	5.2	0.6	5.8	4.9	1.9	7.0	0.2640	42%
Interaction	5.9	3.0	6.8	6.3	1.3	7.0	0.2367	41%
Program Structure	3.8	2.0	4.5	4.5	1.5	5.8	0.0717	32%
Parents and Staff	6.8	6.5	7.0	6.5	3.5	7.0	0.0405*	72%

Note: CL_{QE} is the probability a randomly selected QE classroom will have a higher FCCERS subscale score than a randomly selected MOE classroom. Data are for all FCC homes receiving FCCERS observations across all school years.
* *p*<.05, ***p*<.01, ****p*<.001

Classroom Assessment Scoring System®

In Table D16, Mann-Whitney *U* tests comparing ranked scores across classroom quality levels for spring 2014–15 detected statistically significant differences in ranked Instructional Support Domain scores as well as in all ranked dimension scores under Instructional Support. CL_{QE} for this test indicates a 63 percent probability an MOE classroom would have a higher Instructional Support score than a QE classroom. CL_{QE} for Concept Development indicates a 67 percent probability an MOE classroom would have a higher score than a QE classroom. MOE classrooms also ranked higher than QE classrooms on the Positive Climate dimension of Emotional Support. CL_{QE} for this test indicates a 58 percent probability an MOE classroom would have a higher score than a QE classroom.

Table D16. Comparison of CLASS® Pre-K Domain and Dimension Scores by Classroom Quality Level: Spring 2014–15

	QE (N = 81)			MOE (N = 105)			Mann-Whitney *U* test	
	Median	Min	Max	Median	Min	Max	*p*-Value	CL_{QE}
Emotional Support	6.1	4.5	6.9	6.2	5.2	7.0	0.0772	44%
Positive Climate	6.0	4.5	7.0	6.3	4.5	7.0	0.0242*	42%
Negative Climate	7.0	5.5	7.0	7.0	1.0	7.0	0.2744	52%
Teacher Sensitivity	5.5	3.3	7.0	6.0	3.3	7.0	0.0594	43%
Regard for Student Perspectives	5.5	3.8	6.8	5.5	4.0	7.0	0.1560	46%
Classroom Organization	5.7	3.7	6.8	5.7	4.3	6.7	0.2289	47%
Behavior Management	6.0	3.8	7.0	6.0	3.5	7.0	0.4098	49%
Productivity	6.0	3.3	7.0	6.3	4.3	7.0	0.3799	49%
Instructional Learning Formats	4.8	3.0	7.0	5.3	3.3	6.5	0.0779	44%
Instructional Support	2.9	1.7	5.4	3.2	1.8	5.8	0.0008**	37%
Concept Development	2.3	1.0	5.0	2.8	1.3	6.3	<0.0001***	33%
Quality of Feedback	2.8	1.5	5.8	3.0	1.5	6.3	0.0139**	41%
Language Modeling	3.5	1.8	6.0	4.0	2.3	6.0	0.0007***	44%

Note: results are for evaluation classrooms observed in spring 2014–15.
* $p<.05$, ** $p<.01$, *** $p<.001$

Table D17 shows statistical tests on these data were unable to detect differences in CLASS® Pre-K domain scores over the life of the program. Table D18 also shows for these combined data, Mann-Whitney *U* tests detected no statistically significant differences in ranked CLASS Pre-K domain scores across the three program years.

Table D17. Summary of Findings Across Program Years for Analyses of Differences in Classroom Quality by CLASS® Pre-K Dimension

	2012–13	2013–14	2014–15	2012 –15 Pooled
Emotional Support	No Difference	No Difference	No Difference	No Difference
Classroom Organization	No Difference	No Difference	No Difference	No Difference
Instructional Support	QE < MOE	No Difference	QE < MOE	No Difference

Note: Notable differences between quality levels are associated with Cohen's *d* effect sizes of .2 (small) or larger regardless of statistical significance.

Table D18. Comparison of CLASS® Pre-K Domain Scores by Classroom Quality Level: Three School Years (2012–2015)

	QE (N = 301)			MOE (N = 318)			Mann-Whitney *U* test	
	Median	Min	Max	Median	Min	Max	*p*-Value	CL_{QE}
Emotional Support	6.1	3.7	7.0	6.1	3.8	7.0	0.4446	47%
Classroom Organization	5.7	2.9	7.0	5.7	3.3	6.8	0.2605	51%
Instructional Support	2.9	1.2	5.6	3.0	1.1	6.1	0.0803	47%

Table D19 shows ranked dimension and domain scores were not statistically different across classroom quality levels for 2014–15.

Table D19. Comparison of CLASS® Toddler Domain and Dimension Scores by Evaluation Classroom Quality Level: Spring 2014–15

	QE (N = 22)			MOE (N = 20)			Mann-Whitney *U* test	
	Median	Min	Max	Median	Min	Max	*p*-Value	CL_{QE}
Emotional and Behavioral Support	6.5	4.2	6.9	6.3	4.4	6.9	0.4498	49%
Positive Climate	6.6	5.0	7.0	6.8	3.8	7.0	0.1798	42%
Negative Climate	7.0	1.0	7.0	7.0	5.0	7.0	0.3974	52%
Teacher Sensitivity	6.5	4.0	7.0	6.5	4.5	7.0	0.4645	51%
Regard for Child Perspectives	6.1	4.5	7.0	5.9	3.5	7.0	0.2314	57%
Behavior Guidance	6.0	4.3	7.0	6.0	3.0	7.0	0.2104	43%
Engaged Support for Learning	4.0	2.8	5.3	3.9	1.5	5.3	0.3859	53%
Facilitation of Learning and Development	4.4	3.3	5.8	4.4	1.8	5.8	0.3997	48%
Quality of Feedback	3.8	2.5	5.0	3.4	1.0	5.0	0.3232	54%
Language Modeling	4.0	2.8	5.5	4.0	1.8	5.5	0.4196	52%

CLASS Pre-K Scores and Teacher Education

Table D20 shows CLASS Instructional Support Domain scores were positively associated with the pooled number of ECE or CD units held by classroom teaching staff in 2014–15, regardless of classroom quality level (although the correlation is weak). Table D21 presents similar data using the combined 3-year dataset.

Table D20. Relationship Between Pooled ECE or CD Units Held by Teaching Staff in the Classroom and CLASS® Domain Scores by Classroom Quality Level: 2014–15

	Domain	Spearman's ρ_a	*p*-value
QE	Emotional Support	0.10	0.3251
(n = 104)	Classroom Organization	0.06	0.5299
	Instructional Support	0.26	0.0072**
MOE	Emotional Support	0.26	0.0098**
(N = 100)	Classroom Organization	0.10	0.3166
	Instructional Support	0.21	0.0371*
All	Emotional Support	0.17	0.0129*
(N = 204)	Classroom Organization	0.08	0.2605
	Instructional Support	0.22	0.0014**

Note: Data are for evaluation classrooms, observed in spring 2015, and submitting complete data on ECE units and detailed CLASS scores.
a. Spearman coefficients indicate weak correlations (i.e., <.30).
* *p*<.05, **p*<.01

Table D21. Relationship Between Pooled ECE or CD Units Held by Teaching Staff in the Classroom and CLASS® Domain Scores by Classroom Quality Level: Three School Years (2012–2015)

	Domain	Spearman's ρ_a	*p*-value
QE	Emotional Support	-0.02	0.7892
(n = 289)	Classroom Organization	-0.01	0.8588
	Instructional Support	0.14	0.0209*
MOE	Emotional Support	0.14	0.0122*
(N = 302)	Classroom Organization	0.06	0.3077
	Instructional Support	0.11	0.0505
All	Emotional Support	0.06	0.1231
(N = 591)	Classroom Organization	0.02	0.5541
	Instructional Support	0.11	0.0067**

Note: Data are for evaluation classrooms, observed in spring reporting cycles, across all school years.
a. Spearman coefficients indicate weak correlations (i.e., <.30)
* *p*<.05, **p*<.01

Parent Engagement and Support

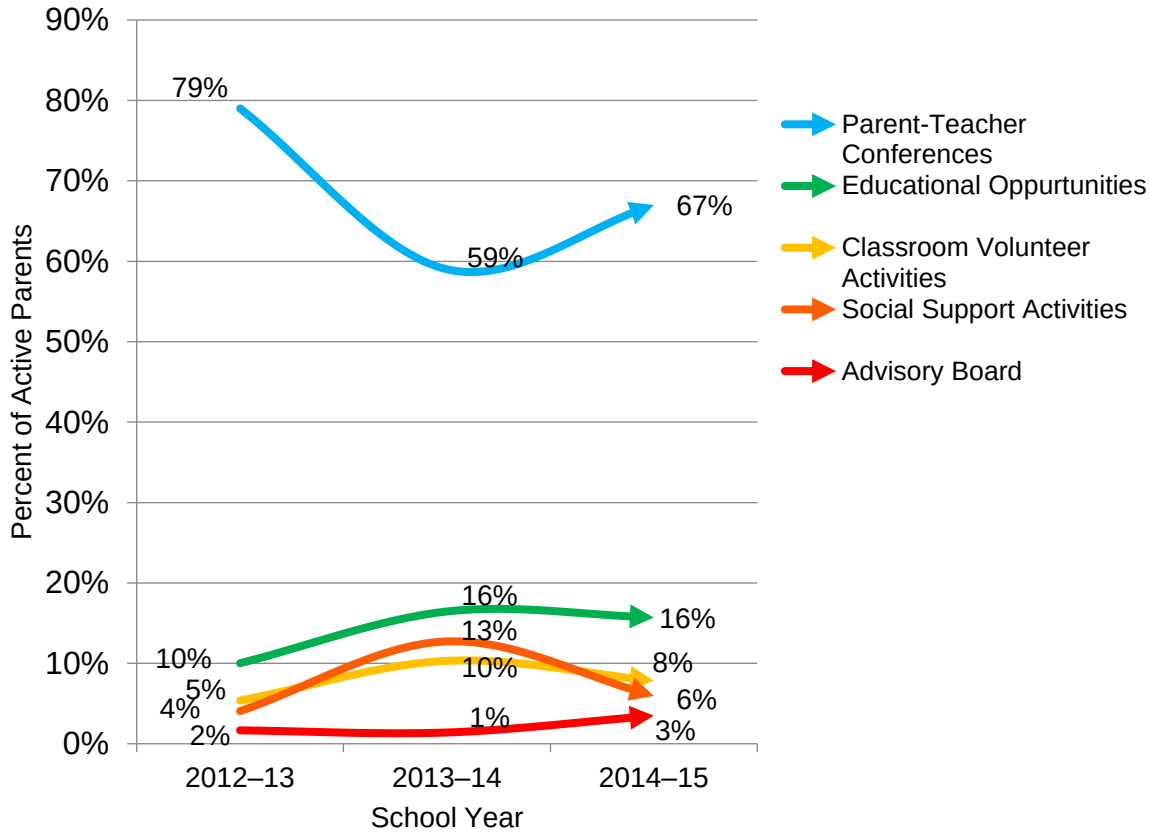
Table D22. Parent Participation by Engagement or Support Activity: 2014–15

Parent Engagement and Support Activity Type	Total Parents Participating	Percent of Active Parents Participating	Parents Participating Per CSP Classroom (*N* = 1,350)	Estimated Percent of Children With a Participating Parent (*N* = 23,640)
Parent-Teacher Conferences	10,568	67%	8	45%
Educational Opportunities	2,477	16%	2	10%
Classroom Volunteer Activities	1,236	8%	1	5%
Social Support Activities	939	6%	1	4%
Advisory Board	551	3%	<1	2%
All Parent Engagement and Support Activities	15,771[a]	100%	12	67%

Note: Active parents are parents who have participated in one or more parent engagement activities. Parents who are more active may participate across multiple engagement and support activities and may be duplicated in this total. Additionally, parents may have multiple children enrolled at the site, and some of these children may or may not be in CSP classrooms.
a. N = 15,771 active parents.

These results should be interpreted with caution because the number of total active parents fluctuated greatly over the life of the program from 21,303 active parents in 2012–13, to a high of 31,823 active parents in 2013–14, and back to 15,771 active parents in 2014–15. These dramatic shifts in parent participation rates are likely related to the 2013 federal sequestration and the drop in Head Start funded classrooms. Additional analyses indicated statistically significant, yet weak, correlations between parent participation rates and the number of Head Start-funded classrooms at a site in 2013–14. And, since Head Start incorporates a strong family engagement component through its *Parent, Family, and Community Engagement Framework* (see Head Start 2011), it is reasonable to suggest that the drop in Head Start-funded classrooms also could have impacted family engagement activities at CSP sites, but this factor alone probably does not fully explain fluctuations in parent participation.

Figure D15. Change in Parent Participation by Engagement or Support Activity Across School Years



Child Development and School Readiness

Figure D16. Child Screening and Assessment: Three School Years (2012–2015)

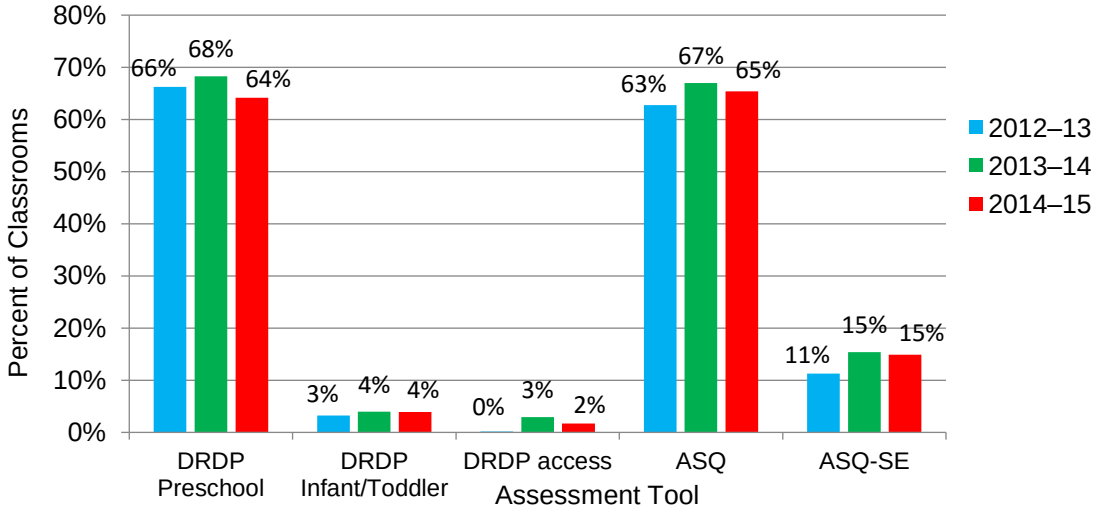


Figure D17. Developmental Screening Results Across School Years

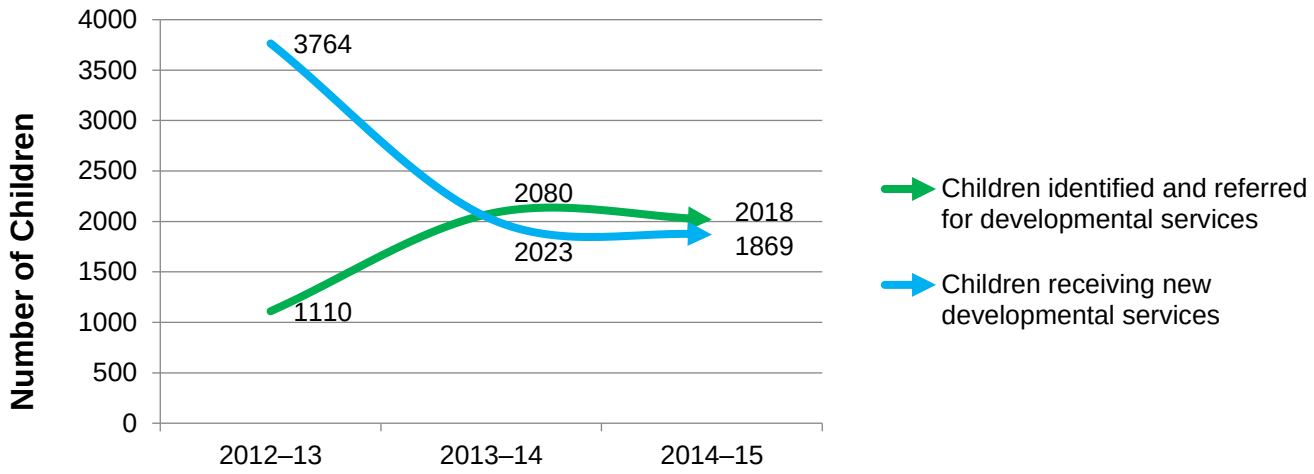


Table D23. Percent of Ratings at the Top Two DRDP-PS Developmental Levels at Fall and Spring by Evaluation Classroom Quality Level: Three School Years (2012–2015)

		Percent Ratings a Top Two Developmental Levels			Difference in Percents (QE – MOE)		*N* Ratings	
Developmental Domain	Classroom Type	Fall	Spring	Gain	Fall	Spring	Fall	Spring
Self and Social Development	QE	35%	83%	+48%	3%***	7%***	123,928	128,964
	MOE	31%	76%	+45%			66,990	70,460
Language and Literacy Development	QE	27%	77%	+50%	2%***	7%***	103,196	106,421
	MOE	25%	69%	+44%			55,712	58,686
English Language Development	QE	38%	76%	+38%	8%***	10%***	30,130	31,592
	MOE	30%	66%	+36%			9,711	15,470
Cognitive Development	QE	34%	82%	+52%	4%***	8%***	51,458	53,223
	MOE	29%	74%	+45%			27,741	29,215
Mathematical Development	QE	27%	78%	+51%	3%***	8%***	62,294	64,086
	MOE	24%	70%	+46%			33,187	35,032
Physical Development	QE	61%	94%	+33%	6%***	7%***	30,902	32,007
	MOE	55%	86%	+31%			16,624	17,510
Health	QE	42%	87%	+45%	2%***	7%***	31,133	31,878
	MOE	40%	80%	+40%			16,751	17,495

Note: N = number of *ratings*, not children. Some DRDP dimensions have more ratings because those dimensions have more measures.
Proportions test significance levels: * $p<.05$, ** $p<.01$, *** $p<.001$. Inconsistencies in differences in percentages are due to rounding.

Table D24. Percent of Ratings at the Top Two DRDP-IT Developmental Levels at Fall and Spring by Evaluation Classroom Quality level: Three School Years (2012–2015)

		Percent Ratings At Top Two Developmental Levels			Difference in Percents (QE – MOE)		*N* Ratings	
Developmental Domain	Classroom Type	Fall	Spring	Gain	Fall	Spring	Fall	Spring
Self and Social Development	QE	32%	48%	+16%	-17%***	-12%***	2,084	1,868
	MOE	49%	60%	+11%			2,000	1,712
Language and Literacy Development	QE	21%	27%	+6%	-16%***	-19%***	962	849
	MOE	37%	47%	+10%			9,060	742
Cognitive Development	QE	36%	48%	+12%	-12%***	-10%***	1,664	1,562
	MOE	48%	58%	+10%			1,663	1,456
Motor and Perceptual Development	QE	49%	52%	+3%	-10%***	-18%***	657	584
	MOE	59%	70%	+11%			584	498
Health	QE	46%	59%	+13%	-18%**	-15%*	157	147
	MOE	64%	73%	+9%			149	132

Note: N = number of *ratings*, not children. Some DRDP dimensions have more ratings because those dimensions have more measures.
* $p<.05$. Inconsistencies in differences in percentages are due to rounding.

The following analyses supplement Tables 8 and 9 (pg. 28 and 29) in the report, and Tables D23 and D24 (pg. 67 and 68) using Cliff's delta effect size as a measure of child development[12]. (See Appendix C in First 5 California 2015 for a fuller description of the utility of Cliff's delta for these analyses.) Table D25 for 2014–15 DRDP-PS data on the following page lists mean Cliff's Delta effect sizes, standard deviations, group size by classroom quality level, difference in effect sizes between classroom quality levels, Mann-Whitney *U* test results, and CL_{QE} effect sizes for each developmental domain. *U* tests did not detect statistically significant differences in ranked effect sizes between classroom quality levels. Additionally, with the exception of English Language Development for DLLs, differences in mean effect sizes were negligible (yet in the expected direction). CL_{QE} for English Language Development indicates a 55 percent probability a randomly selected MOE classroom would have a higher effect size than a randomly selected QE classroom. As in prior school years, consistent medium and large effect sizes (0.36 to 0.63) across classroom quality levels and developmental domains

[12] Cliff's delta is an effect size measure quantifying how much the distributions of fall and spring DRDP ratings diverge or overlap (see Cliff 1996 and First 5 California 2012b). A zero represents complete overlap (i.e., the distributions are not different) and a 1 or -1 indicates perfect divergence (i.e., distributions are completely different). Deltas of 0.147, 0.33, and 0.474 correspond with Cohen's *d* effect sizes of 0.2 (small), 0.5 (medium), and 0.8 (large) (Cohen 1988 and Romano et al. 2006). Cliff's delta accounts for where each child starts along the DRDP developmental continuum, how that child's position relates to the positons of other children in the classroom, and works to produce a relative effect size for each classroom and each domain.

suggest that children experienced healthy development in CSP regardless of classroom quality level in 2014–15.

Table D26 presents the same information using DRDP-PS data from all school years, 2012–15. For these combined data, *U* tests detected statistically significant differences in ranked effect sizes between classroom quality levels for Mathematical Development and Physical Development at the $p<.05$ level. Additionally, significance levels for Self and social Development, Language and Literacy Development, English Language Development, and Cognitive Development approached statistical significance at the $p<.10$ level. CL_{QE} for Physical Development indicates a 55 percent probability a QE classroom would have a higher effect size than an MOE classroom (a small effect), and CL_{QE} for Mathematical Development indicates a 53 percent probability a QE classroom would have the higher effect size. Consistent medium and large Cliff's delta effect sizes (0.46 to 0.63) across all DRDP-PS developmental domains, for both QE and MOE classrooms, suggest children experienced healthy development over the life of the program regardless of classroom quality level.

Table D25. Results of Analyses of Differences in Effect Sizes (*d*) Across Classroom Quality Levels by DRDP-PS Developmental Domains: 2014–15

	QE			MOE					
DRDP Domain	Fall-to-Spring Mean Cliff's *d* Effect size	SD	*N*	Fall-to-Spring Mean Cliff's *d* Effect Size[a]	SD	*N*	Difference in Mean Effect Size[b]	Mann-Whitney *U* *p*-Value	CL_{QE}
Self and Social Development	0.60 (large)	0.25	69	0.59 (large)	0.29	73	0.01	0.42	49%
Language and Literacy Development	0.61 (large)	0.23	68	0.57 (large)	0.28	73	0.04	0.22	54%
English Language Development	0.36 (medium)	0.51	67	0.51 (large)	0.30	71	-0.15	0.17	45%
Cognitive Development	0.62 (large)	0.26	68	0.57 (large)	0.31	73	0.05	0.28	53%
Mathematical Development	0.63 (large)	0.24	68	0.59 (large)	0.30	73	0.04	0.42	51%
Physical Development	0.57 (large)	0.29	68	0.55 (large)	0.34	73	0.01	0.49	50%
Health	0.45 (medium)	0.30	67	0.44 (medium)	0.32	73	0.01	0.48	50%

Note: CL_{QE} is the probability a randomly selected QE classroom will have a higher mean Cliff's *d* effect size than a randomly selected MOE classroom. Data are for all evaluation classrooms submitting DRDP pre and post results across all school years.

a. Cliff's Delta effect sizes of 0.147 (small), 0.33 (medium), and 0.474 (large) correspond to Cohen's *d* effect sizes of 0.2 (small), 0.5 (medium), and 0.8 (large).

b. Negative differences in mean effect sizes indicate MOE classrooms outperformed QE for the developmental domain.

Table D26. Results of Analyses of Differences in Effect Sizes (*d*) Across Classroom Quality Levels by DRDP-PS Developmental Domains: Three School Years (2012–2015)

	QE			MOE					
DRDP Domain	Fall-to-Spring Mean Cliff's *d* Effect size	SD	*N*	Fall-to-Spring Mean Cliff's *d* Effect Size[a]	SD	*N*	Difference in Mean Effect Size[b]	Mann-Whitney *U* *p*-Value	CL_{QE}
Self and Social Development	0.61 (large)	0.32	618	0.59 (large)	0.31	326	0.02	0.09	53%
Language and Literacy Development	0.60 (large)	0.33	612	0.58 (large)	0.30	325	0.02	0.07	53%
English Language Development	0.46 (medium)	0.42	585	0.48 (large)	0.31	316	-0.02	0.06	53%
Cognitive Development	0.61 (large)	0.35	609	0.59 (large)	0.33	323	0.02	0.06	53%
Mathematical Development	0.63 (large)	0.33	610	0.60 (large)	0.32	323	0.03	0.04*	53%
Physical Development	0.58 (large)	0.34	610	0.53 (large)	0.36	321	0.05	0.01**	55%
Health	0.49 (large)	0.35	607	0.46 (medium)	0.36	321	0.03	0.12	52%

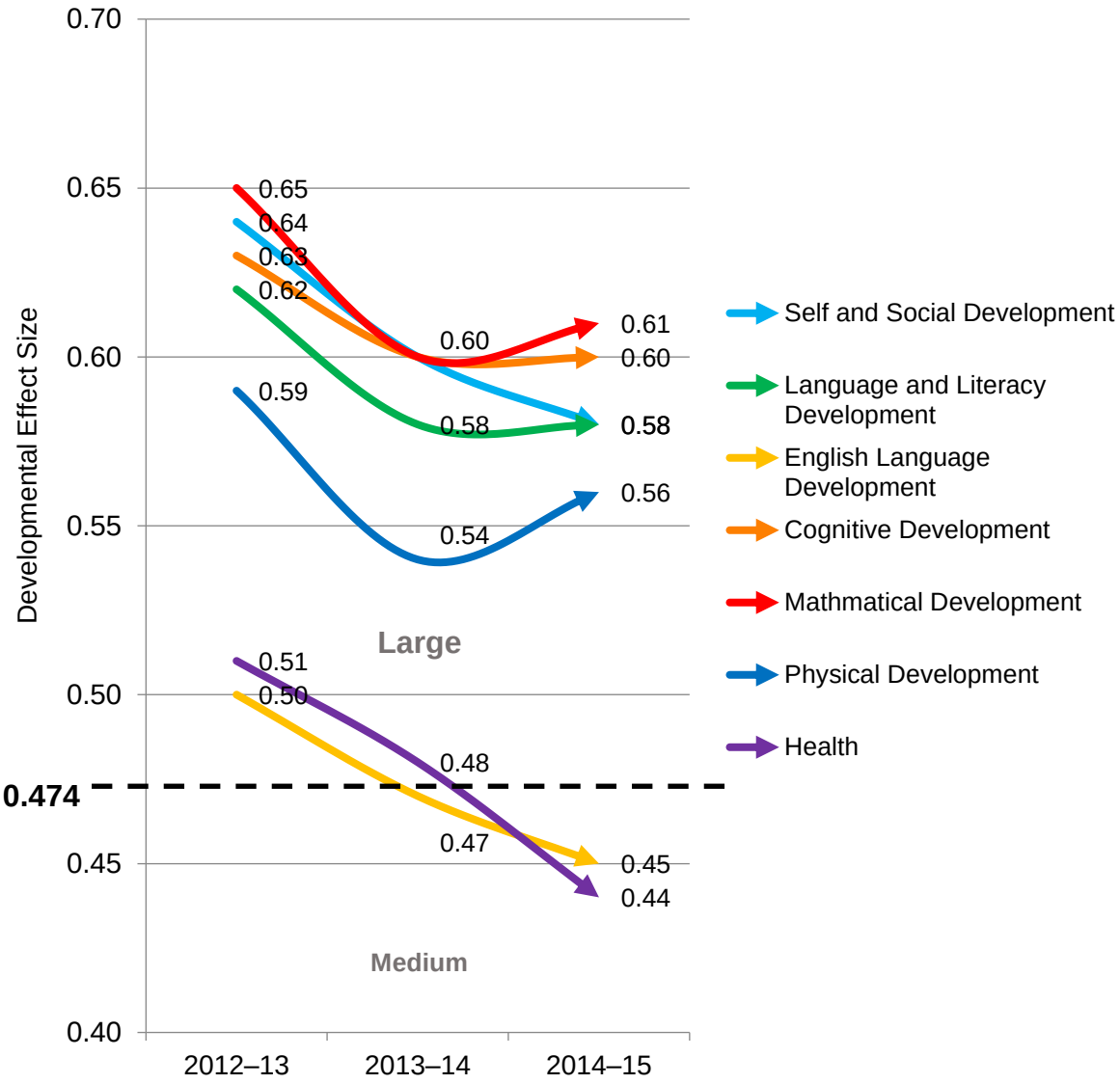
Note: CL_{QE} is the probability a randomly selected QE classroom will have a higher mean Cliff's *d* effect size than a randomly selected MOE classroom. Data are for all evaluation classrooms submitting DRDP pre and post results across all school years.

a. Cliff's Delta effect sizes of 0.147 (small), 0.33 (medium), and 0.474 (large) correspond to Cohen's *d* effect sizes of 0.2 (small), 0.5 (medium), and 0.8 (large).

b. Negative differences in mean effect sizes indicate MOE classrooms outperformed QE for the developmental domain.

Figure D18 depicts change in mean Cliff's delta effect sizes for all DRDP-PS developmental domains and for all classrooms submitting fall and spring DRDP-PS results across school years. These data also show consistent medium and large (>0.33) developmental effect sizes over the life of the program and suggest preschool children experienced healthy development in CSP Whether in QE or MOE classrooms.

Figure D18. Mean Developmental Effect Sizes by DRDP-PS Domain Across QE and MOE Pooled Ratings, 2012–2015



Note: the horizontal dotted line at effect size 0.474 indicates the threshold differentiating medium from large effects. Cliff's *d* Effect sizes above 0.474 are considered large effects.

Table D27 (pg. 74) for DRDP-IT 2012–15 data lists mean Cliff's Delta effect sizes, standard deviations, group size by classroom quality level, difference in effect sizes between classroom quality levels, Mann-Whitney *U* test results, and CL_{QE} effect sizes

for each developmental domain. For these combined data, *U* tests did not detect any statistically significant differences in developmental effects between QE and MOE classrooms. With the exception of Language and Literacy Development in MOE classrooms, developmental effect sizes were small. Although not statistically significant, the difference in mean developmental effect size for Language and Literacy Development was a medium-sized effect. CL_{QE} for this result indicates a 58 percent probability an MOE classroom would have a higher effect size than a QE classroom. In prior reports, it was not possible to make meaningful comparisons of QE and MOE infant/toddler classrooms in terms of developmental effects because sample sizes were small.

Table D27. Results of Analyses of Differences in Effect Sizes (*d*) Across Classroom Quality Levels by DRDP-IT Developmental Domains: Three School Years (2012–2015)

	QE			MOE					
DRDP Domain	Fall-to-Spring Mean Cliff's *d* Effect size	SD	*N*	Fall-to-Spring Mean Cliff's *d* Effect Size[a]	SD	*N*	Difference in Mean Effect Size[b]	Mann-Whitney *U* *p*-Value	CL_{QE}
Self and Social Development	0.21 (small)	0.23	19	0.25 (small)	0.26	19	-0.04	0.34	46%
Language and Literacy Development	0.20 (small)	0.28	19	0.53 (large)	1.39	20	-0.33	0.20	42%
Cognitive Development	0.20 (small)	0.23	19	0.25 (small)	0.30	19	-0.05	0.43	48%
Motor and Perceptual Development	0.18 (small)	0.29	19	0.28 (small)	0.31	19	-0.10	0.15	40%
Health	0.15 (small)	0.45	19	0.20 (small)	0.32	18	-0.05	0.40	50%

Note: CL_{QE} is the probability a randomly selected QE classroom will have a higher mean Cliff's *d* effect size than a randomly selected MOE classroom.

a. Cliff's Delta effect sizes of 0.147 (small), 0.33 (medium), and 0.474 (large) correspond to Cohen's *d* effect sizes of 0.2 (small), 0.5 (medium), and 0.8 (large).

b. Negative differences in mean effect sizes indicate MOE classrooms outperformed QE for the developmental domain.