THIRD SECTOR
INTELLIGENCE INC.

NIEER
NATIONAL INSTITUTE FOR
EARLY EDUCATION RESEARCH

Seattle Preschool Program

Recommended Comprehensive Evaluation Strategy

7-30-2015

Contributors

- Third Sector Intelligence (3SI) Team
    - Chris Strausz-Clark, MA, Managing Principal
    - Maria Gingerich, MA, Senior Manager
    - Joelle Gruber, MA, Consultant
- National Institute for Early Education Research (NIEER) Team
    - W. Steven Barnett, PhD, Director
    - Milagros Nores, PhD, Associate Director of Research
    - Allison Friedman-Krauss, Assistant Research Professor
    - Jessica Francis, MEd, Research Fellow
- Augenblick Palaich and Associates (APA) Team
    - Anne Mitchell, MA, Consultant
    - Simon Workman, MA, Associate
- Advisors
    - Kathleen Bruck, Chief Executive Officer, PRE-K 4 SA
    - Gail Joseph, PhD, Associate Professor, University of Washington and Director, Childcare Quality and Early Learning Center for Research and Professional Development, University of Washington
    - Jim Minervino, Senior Advisor to 3SI and Founder and CEO, Ready On Day One
    - Phil Sirinides, PhD, Senior Advisor to 3SI and Senior Researcher, Consortium for Policy Research in Education, University of Pennsylvania
    - Valisa Smith, MA, Senior Advisor to 3SI

Special thanks to:

- Ellen Frede, PhD, Deputy Director, Early Learning, The Bill and Melinda Gates Foundation and former Assistant to the Commissioner for Early Childhood Education at the New Jersey Department of Education
- Jason Sachs, PhD, Director, Early Childhood Education, Boston Public Schools

Contents

Glossary and Acronyms

Term	Definition[1]
Bias	Bias is a form of systematic error that can affect scientific investigations and distort the measurement process. A biased study loses validity with respect to the degree of the bias. One type of research bias, selection bias, which may result in the subjects in the sample being unrepresentative of the population of interest.[2]
Classroom Assessment Scoring System (CLASS)	CLASS is an observational tool used to assess classroom practices by measuring the interactions between students and teachers.
Comparison Group	A comparison group is a group of students not attending the SPP but virtually identical to the group that attended SPP.
Continuous Quality Improvement (CQI)	A process-based, data-driven approach to improving the quality of a product or service through standards, measurement, analysis, and improvement.
Cultural Competency	Set of congruent behaviors, attitudes, and policies that come together in a system, agency, or profession that enables that system, agency, or profession to work effectively in cross-cultural situations.[3]
Culturally Responsive Education	Using cultural knowledge, prior experiences, frames of reference, and performance styles of ethnically diverse students to make learning more relevant and effective for them.[4]
DEEL Coaches	Full-time DEEL staff members on the Quality Assurance Team focused on teachers' professional growth through on-site coaching and training.
DEEL Education Specialists	Full-time DEEL staff members on the Operations Team who monitor compliance with SPP performance standards and provide technical assistance to providers.
DEEL Operations Team	Oversees student enrollment, administration of preschool assignment process, compliance with SPP program standards, fiscal/technical oversight, and program scopes of work.
DEEL Policy and Planning Team	Oversees special projects (including but not limited to the evaluation strategy), coordination with state and regional efforts, grant writing, and legislative coordination.
DEEL Quality Assurance Team	Oversees coaching, training and professional development, site assessments, and curriculum.
Dual Language Learners (DLL)	Children who are Dual Language Learners acquire two or more languages simultaneously, learning a second language while continuing to develop their first language.
Early Achievers (EA)	Washington State's quality rating and improvement system.

[1] Where relevant, sources are available when terms are used in the body of this document.
[2] Krishna, R., Maithreyi, R., Surapaneni, K. M. (2010). Research Bias: A Review For Medical Students. Journal of Clinical and Diagnostic Research, (4),2320-2324.
[3] Cross T., Bazron, B., Dennis, K., & Isaacs, M. (1989). Towards a culturally competent system of care, volume I. Washington, D.C.: Georgetown University Child Development Center, CASSP Technical Assistance Center.
[4] Gay, Geneva. (2002). Preparing for Culturally Responsive Teaching. Journal of Teacher Education, 53,106.

Term	Definition[1]
Early Childhood Education and Assistance Program (ECEAP)	Washington's state-funded comprehensive preschool program that provides free services and support to eligible children and their families.
Early Childhood Environment Rating Scale (ECERS-3/ECERS-R)	ECERS-3 is a classroom observational tool used to evaluate student experiences that have a direct effect on development, including interactions between teachers and students and the interactions students have with the classroom environment. The ECERS-3 builds upon the foundation provided by the earlier ECERS-R.
Explicit Bias	Explicit bias refers to attitudes that are consciously controlled and reflect what people are willing and able to admit to themselves and others.[5]
Family Child Care (FCC)	Family Child Care (FCC) is a service wherein child care is provided in a caregiver's own home.
Fidelity of Implementation	The degree to which a program is delivered as intended.
Head Start Child Development and Early Learning Framework	The Head Start Child Development and Early Learning Framework provides Head Start and other early childhood programs with a description of the developmental building blocks that are most important for a child's school and long-term success.
Impact Evaluation	Outcome-focused measurement of the effect of a program.
Implicit Bias	Refers to the attitudes or stereotypes that affect our understanding, actions, and decisions in an unconscious manner. These biases, which encompass both favorable and unfavorable assessments, are activated involuntarily and without an individual's awareness or intentional control.[6]
Integrated Systems	Systems that support the combined use of targeted professional development, use of data, curriculum, standards, and assessments.
Managed Education and Registry Information Tool (MERIT)	MERIT is an online tool used to document and recognize the professional achievements of early care and education and school-age professionals in the state of Washington.
Mixed-Delivery Model	Public schools and community-based organizations provide preschool services.
Pre-Post Design	A pre-post study examines whether program participants demonstrate improved outcomes over a set time period (such as a year) by measuring at program entry and exit.
PRISM	Data management system organizing EA ratings data.
Process Evaluation	Determines whether the program has been implemented as intended.
Qualitative Data	Qualitative data describe the attributes or properties that an object possesses.
Quality Improvement Plan (QIP)	A plan for improvement, developed by teachers and coaches.
Quality Rating and Improvement System (QRIS)	A program to assess, improve, and communicate the level of quality in early care and education settings.

[5] Seattle DEEL.
[6] Implicit Bias Review: http://kirwaninstitute.osu.edu/wp-content/uploads/2014/03/2014-implicit-bias.pdf

Term	Definition[1]
Quantitative Data	Quantitative data express a certain quantity, amount, or range.
Race and Social Justice Initiative (RSJI)	Seattle's commitment to eliminate racial disparities and achieve racial equity in the city.
Randomized Controlled Trial (RCT)	A study design that randomly assigns participants into an experimental group or a control group.
Regression Discontinuity Design (RDD)	A study design that assigns a cutoff or threshold for determining students' eligibility for the program.
Relational Data Management System	Provides the ability to use tables for data storage while maintaining and enforcing certain data relationships.
Reliability	The extent to which an experiment, test, or measuring procedure yields the same results on repeated trials.
Selection Bias	Arises when participants in a program are systematically different from nonparticipants (even before they enter the program). Many evaluations compare program participants to nonparticipants in order to infer the effect of the program; selection bias can affect the legitimacy of these evaluations, and, in particular, we believe that its presence is likely to skew evaluations of nonprofits in the positive direction.[7]
Self-Assessment Validation System (SAVS)	A rubric completed by providers about NJ program standards as part of a multiphase process of program improvement.
Special Populations	Children who are in the child welfare system, live in transitional housing, come from homeless families, have healthcare needs, have mental health needs (as evidenced by behavioral screening and parent/teacher/coach observation), or have developmental needs (as evidenced by developmental screening and parent/teacher/coach observation).
SPP Program Standards	The core components of SPP as outlined in the Action Plan proposed by Mayor Murray and amended by the Seattle City Council.
Step Ahead	Seattle-funded full- and half-day high-quality preschool programs for low-income 3- and 4-year-olds.
The Essential Elements of High Quality Pre-K	The 15 common practices of four public preschool programs that have high-quality outcomes for students.
TS Gold	Teaching Strategies GOLD is an authentic, ongoing observational system for assessing children from birth through Kindergarten. It helps teachers to observe children in the context of everyday experiences, which is an effective way to learn what they know and can do.
Utilization-focused	A decision-making framework for enhancing the utility and actual use of evaluations.
Validity	The extent to which an experiment, test, or measuring procedure measures what it is purported to measure.
Web-based Early Learning System (WELS)	The QRIS database that tracks rating information, quality improvement plans, and coaching data.

[7] Common Problems with Formal Evaluations: Selection Bias and Publication Bias: http://www.givewell.org/united-states/process/common-evaluation-problems

Section I: Introduction

The Role of Evaluation in the Seattle Preschool Program (SPP)

On November 4, 2014, Seattle voters approved a four-year, $58 million property tax levy to provide "accessible high-quality preschool services for Seattle children designed to improve their readiness for school and to support their subsequent academic achievement."[8]

The city is investing SPP levy proceeds to achieve the following outcomes city-wide:

- Children will be ready for school.
- All students will achieve developmentally appropriate pre-academic skills.
- All students will develop both socially and emotionally.
- The readiness gap will be eliminated for SPP participants.

The city of Seattle's Department of Education and Early Learning (DEEL) will launch SPP in the 2015–16 school year and expand it rapidly over the next three years. Table 1 shows estimated targets for number of classrooms and children served. (Note that these are estimates and not fixed targets.)

Table 1: Estimated targets for SPP classrooms and children served

School Year	Targeted Number of Classrooms	Estimated Number of Children Served
2015–16	14 classrooms	280 children
2016–17	39 classrooms	780 children
2017–18	70 classrooms	1,400 children
2018–19	100 classrooms	2,000 children

The four-year demonstration phase of SPP has three purposes. The first is to ascertain proof of concept. In other words, SPP must demonstrate that the approved structure is viable and has the capacity to produce positive outcomes for Seattle's children.

The second purpose is to create, refine, and support a community infrastructure to improve the quality of preschool programs. Specifically, over the next six months, DEEL will:

[8] City of Seattle Proposition 1B, preamble.

HOW THIS DOCUMENT IS ORGANIZED

This document is organized according to the key activities in the CQI cycle:

- Operationalize Standards and Expectations
- Measure and Collect Data
- Analyze Results and Plan
- Implement improvements

Additional sections of the document address specific topics, such as the development of the Family Child Care pilot and evaluation, and the projected four- year budget.

Each of the four sections of this document that relate to the CQI cycle contain an introduction that explains the step in the cycle and why it is important to the evaluation strategy; a brief synthesis of results of the literature review that informed the SPP Evaluation Strategy; and the recommended approach for SPP including tools, evaluators, and timing.

- Design a process through which community preschool providers can access funding to improve, expand, and renovate facilities to provide additional classroom space for SPP.
- Work with local community colleges to ensure higher education programs are accessible and responsive to the needs of the early learning workforce.
- Work with school district and state partners to align systems and leverage resources in a non-duplicative manner.
- Work with community partners to create a site-level assessment rubric that accurately assesses providers' progress toward quality standards and identifies areas of needed support.
- Support providers in achieving the high standards of SPP through embedded coaching and professional development.
- Identify practical approaches to blending and braiding disparate publicly funded preschool programs (such as ECEAP and Head Start).

Addressing these challenges and achieving quality takes time, and the research on high-quality early learning programs across the country reinforces this.

The third purpose of the demonstration phase of SPP is to create a process and norms that support continuous quality improvement (CQI) through evaluation. Like preschool programs that have demonstrated effectiveness, Seattle will use results from its initial years of evaluation to make course corrections to its programs. For example:

- After focusing the first two years of its evaluation on measuring classroom quality, including assessing the effectiveness of curricula, Boston decided to implement more effective curricula and provide additional support for teachers.[9]
- New Jersey began by collecting data on program implementation, classroom quality, and child outcomes during the initial years of the Abbott preschool program. This information was used to focus statewide professional development and technical assistance on key areas that needed improvement.[10]

High-quality preschool programs can produce meaningful improvements in children's school readiness and school success. The school readiness improvements most often observed include gains in vocabulary, early literacy skills, mathematics, self-control and attention skills, and social skills. Later in life, children who attended high-quality preschool have improved achievement test scores, are less likely to fail a grade, are less likely to require special education services, and are more likely to graduate from high school on time.

Until race and family income are no longer predictive of school performance, the city has committed to making investments that will help all of Seattle's children succeed in school and life. The city of Seattle is dedicated to ensuring all children have high-quality early learning opportunities and that every aspect of SPP advances racial equity and social justice and aligns with the city's Race and Social Justice Initiative.

[9] Boston's Rapid Expansion of Public School-Based Preschool: https://www.naeyc.org/files/academy/file/YCSept2010.pdf
[10] Interview with Ellen Frede, PhD, 6/17/2015.

"The Seattle Race and Social Justice Initiative (RSJI) is a citywide effort to end institutionalized racism and race-based disparities in City government. RSJI builds on the work of the civil rights movement and the ongoing efforts of individuals and groups in Seattle to confront racism. The Initiative's long term goal is to change the underlying system that creates race-based disparities in our community and to achieve racial equity."

—City of Seattle, 2014

In Seattle today, economic and racial disparities persist in third-grade reading levels, fourth-grade math levels, and high school graduation rates. Disparities linked to family income and race can persist throughout a student's academic career. In Seattle and across Washington State, these deficits were more pronounced for children of color (see the Seattle Preschool Program Plan for more information).[11] Continuously analyzing results will ensure SPP is working toward achieving its goals.

The Continuous Quality Improvement (CQI) Cycle

To ensure SPP successfully adapts to the needs of all Seattle children, it is necessary to collect and analyze data that helps the program adapt and improve. To do this, Seattle will apply a Continuous Quality Improvement (CQI) framework to all aspects of its evaluation, as shown in Figure 1.

Figure 1: Continuous Quality Improvement Cycle for the Seattle Preschool Program

In the CQI model, data are collected and results are analyzed, and improvements are made cyclically. The entire evaluation process is treated as continuous.

A key benefit of the CQI evaluation approach is that it is responsive. Analyzing results will identify areas for improvement and inform hypotheses about why aspects of a program are not working, which may require additional data collection. As such, CQI allows DEEL to strike a balance between adhering to a consistent evaluation strategy and adjusting to emerging needs and changes as SPP expands, evolves, and improves and the landscape of early learning in Washington State evolves.

Using a CQI framework also demands that the SPP evaluation strategy is utilization-focused. A utilization-focused evaluation explicitly engages key stakeholders in designing and executing the

[11] See the Seattle Preschool Program Plan: http://www.seattle.gov/Documents/Departments/OFE/AboutTheLevy/EarlyLearning/SPP_ProgramPlan2015-16_Final.pdf

evaluation. Ultimately, the value of the evaluation will be in the team's ability to draw conclusions from evaluation and identify appropriate course corrections.

Lessons Learned from the Literature Review and Other Programs

To develop SPP's evaluation strategy, the Evaluation Team conducted a thorough review of the research on evaluation, supplemented with interviews of key leaders in program design and improvement. This review focused in particular on studies of large-scale public preschool programs administered by cities and states, including some specifically identified by the city of Seattle as highly relevant to the SPP. This review is available as a companion document and is referenced throughout this strategy.[12] The overarching lessons learned are:

- According to the research, teacher-student interactions are at the core of what is a high-quality experience for children. As such, evaluations should include and focus on teacher-student interactions and the processes that contribute to these interactions.
- Like the children they serve, preschool programs take time to mature, and evaluations of newly created programs must take this into account. The evaluations that found the largest effects on student outcomes are studies of mature programs that have had time to improve.[13] Effects on school readiness and other child outcomes are typically smaller in the first years of a new preschool program.[14]
- Programs fully mature at varying rates. The rate of progress depends on the capacity and quality of programs and staff at the start, the resources provided for change, and the specifics of standards programs, teachers, and others are required to achieve.
- High-quality programs begin evaluation at inception by collecting information on students, classrooms, and practices. These results are used to make continuous improvements.

Definition and Types of Evaluation

According to the American Evaluation Association, evaluation is a systematic process to determine merit, worth, value, or significance.[15] Evaluation can explain how a program affects the participants and help explain why a program works or does not work. Well-conceived evaluations provide information that is important for continuous program improvement and simultaneously serve as a report card on programs to policymakers and the public.

Evaluation can be characterized as either *impact* evaluation or *process* evaluation. Impact evaluation analyzes the short- to long-term outcomes of a program on the target group to demonstrate whether or

[12] Nores, M., et al. (2015). Technical Report for the City of Seattle. A Review of the Evidence on Preschool Programs and a Comparison of Selected State and City Programs.
[13] Ibid.
[14] Minervino, Jim (with contribution from Robert C. Pianta, PhD, University of Virginia). Lessons from Research and the Classroom: Implementing High-Quality Pre-K that Makes a Difference for Young Children (White Paper, Bill & Melinda Gates Foundation, September 2014).
[15] American Evaluation Association: http://www.eval.org/p/bl/et/blogid=2&blogaid=4

not the program is producing the expected results. Process evaluation provides information on how the program could be improved and, over time, tracks what leads to a program working. Process evaluation measures whether the program has been implemented as intended, using a combination of qualitative and quantitative methods, to better understand whether and how the program creates desired impact.[16] The process evaluation analyses can be used to identify what needs to be improved if desired impacts are not achieved. Process evaluation takes place in the early stages of a program implementation and, ideally, throughout the program's lifetime.

SPP's Approach to Evaluation

The Seattle Preschool Program Evaluation Strategy is based on a comprehensive review of the preschool evaluation literature and the Evaluation Team's research on approaches that other cities and states have used to evaluate their preschool programs. (See Appendix A for a summary of approaches taken by other cities and states; see the companion "Technical Report for the City of Seattle" for a more thorough comparison.) The first stage of the evaluation will emphasize measuring and analyzing implementation of the SPP and children's experiences in the SPP. It will produce baseline information on SPP and children's learning and development. This information will allow all stakeholders to better understand how the SPP is working. Once the SPP shows expected degrees of quality and maturation, a second stage will focus on continuing improvements but also capture impact trends on children's Kindergarten readiness.

We will know that the SPP has achieved "maturity" when the majority of providers in the program have achieved most or all of the standards for classroom quality, based on direct observations by experts in early education.

The specific evaluation methods that will be used in the SPP evaluation

The SPP evaluation strategy will include three distinct evaluation methods:

1. An **impact evaluation** to demonstrate to what degree SPP is increasing Kindergarten readiness on the students it serves
2. A **process evaluation** to assess implementation of the SPP program and evaluate quality and consistency amongst SPP preschool providers
3. A **self-evaluation** to help SPP providers, in partnership with and supported by DEEL staff, measure and improve their own programs

There is some overlap and dependency between each of these evaluation methods, but they represent distinct aspects of evaluation, performed at specific times.

[16] Community Sustainability Engagement Evaluation Toolbox: http://evaluationtoolbox.net.au/index.php?option=com_content&view=article&id=24:formative-evaluation&catid=17:formative-evaluation&Itemid=125

Impact evaluation

The Evaluation Team will implement a research-based experimental design to determine if and when SPP achieves its desired impacts on Kindergarten readiness. To do this, the external Evaluation Team will:

- Assess children's development in language, literacy, math, and executive functions[17]
- Assess children's learning, development, and well-being before and after participating in SPP
- Compare the progress of participants to that of nonparticipants to isolate the effects of SPP as an intervention

As outlined in the Measurement section, the impact evaluation will become increasingly robust as SPP moves through the Demonstration Phase (2015–19). Initially the child outcome data will be used to identify areas in need of program improvement. As SPP matures and the sample sizes increase, the impact evaluation will be able to demonstrate progress, or lack thereof, toward increasing Kindergarten readiness. Expectations for impacts on children's learning and development should be adjusted for the program's maturity. As described in the sections below, the impact evaluation will involve stakeholders from the Washington State Department of Early Learning (WA DEL) and center directors and teachers. The results of the impact evaluation will be shared at varying degrees of specificity, at WA DEL's discretion, with center directors and teachers, lawmakers and community leaders. The Evaluation Team will provide analyses of student-level performance to DEEL as described in Section IV (Analysis). Student-level performance improvements will be based on DEEL's ability to solidify infrastructure, implement processes, and scale the program.

Process evaluation

The process evaluation will assesses whether SPP is on track in executing the program model to achieve desired outcomes and measure the extent to which providers are implementing the SPP program with fidelity. The process evaluation will:

- Use measures of the quality of teacher-student interactions, classroom environments, and classroom processes, as well as data from the self-evaluation to evaluate SPP quality
- Leverage data available through the WA DEL data systems on teacher credentials, professional development, and the Quality Rating and Improvement System (QRIS)
- Use select administrative and quality data collected by DEEL to provide additional information on process and structural quality for the programs in SPP
- Gather feedback on the program from directors, teachers, and families

As described in later sections, the process evaluation will involve data collection from DEEL, center directors, and teachers, as well as other SPP partners such as WA DEL. The process evaluation will provide important feedback to DEEL in its early years and allow DEEL to identify and implement improvements.

[17] The SPP evaluation will also collect measures of social and emotional development.

Self-evaluation

As part of self-evaluation, DEEL staff will evaluate the progress of individual providers in implementing SPP program design, help providers assess and identify opportunities for improvements, monitor compliance with SPP standards, and support directors and teachers in implementing improvements.[18] The self-evaluation will allow providers to evaluate and improve their own programs, contribute qualitative and quantitative data about classroom quality to the evaluation, and serve as a mechanism through which the city can scrutinize the quality of its internal processes for the purposes of improvement.

How the three evaluation methods work together

High-quality classroom environments and practices produce comparatively more benefits for children, but classrooms that are of high quality in one respect and weak in other areas will likely produce uneven child outcomes. For example, many preschool classrooms are better at supporting social development than cognitive development or do a better job with language arts than with math. SPP seeks excellence in multiple areas, so it is essential that the evaluation assess classroom quality broadly, using all three approaches to evaluation.

All three methods of evaluation focus on high-quality classrooms and teacher-student interactions. Each of these evaluation methods has a distinct role in the SPP evaluation, although some measurements—such as observations of classroom quality—will contribute to more than one evaluation method. See the next section, Timeline and Roles and Responsibilities for Evaluation, for more details about the three types of evaluation and their interdependence.

For each evaluation method, Table 2 describes who will be responsible for measuring, collecting data, analyzing results, and planning for improvements.

Table 2: Roles and responsibilities for each evaluation component

Evaluation Component	Performed by	General Timing
Impact evaluation	External Evaluation Team (For 2015–16: NIEER and UW)	Annual report in the summer with status updates in the winter and spring
Process evaluation	External Evaluation Team (For 2015–16: 3SI)	Annual report in the summer with interim progress reports in the winter and spring to inform continuous improvement
Self-evaluation	DEEL coaches and education specialists[19]	Ongoing relationships between coaches and preschool providers; Reporting on quarterly cycle[20]

The three evaluation methods will be implemented as continuous and overlapping processes, although each evaluation method will have its own distinct cycle. The impact and process evaluation reporting

[18] DEEL staff includes coaches and education specialists within the Quality Assurance and Operations teams respectively.
[19] DEEL's Coaching Teams include the Quality Assurance Team and education specialists on the Operations Team.
[20] Measurement and reporting frequency will be finalized as DEEL develops the self-evaluation approach.

will fall on an annual cycle, and the self-evaluation will be on a more frequent cycle. Although the specific timeline for evaluation activities will be developed in partnership with DEEL, the general cycle of activities is to collect data or measure over the school year (fall and spring), analyze data, and report findings to DEEL each summer. This will enable DEEL and its oversight bodies to review outcomes regularly and work with program leaders and providers to refine and improve the program prior to the beginning of the new school year. As the school year begins and providers implement improvements, data collection and analyses for the next cycle will begin.

Moving Forward

The city of Seattle is committed to developing the infrastructure and processes necessary to support high-quality child outcomes and meet its goals. The Evaluation Team is committed to supporting the city by collecting and analyzing data, creating reports, and providing evidence-based course corrections and recommendations when needed to help the city meet its goals.

Section II: Align Evaluation Strategy with Program Standards

The key program design elements of the SPP are firmly rooted in evidence-based practice.[21] The city of Seattle, in partnership with BERK Consulting and local and national experts, created SPP's standards based on research, outcomes from early learning work groups, and feedback from key stakeholders such as Seattle Public Schools (SPS), Washington State Department of Early Learning (WA DEL), and community-based organizations.[22]

As stated in Section I, the Seattle Preschool Program Plan identifies the goals of the SPP as:

- Children will be ready for school.
- All students will achieve developmentally appropriate pre-academic skills.
- All students will develop both socially and emotionally.
- The readiness gap will be eliminated for SPP participants.

Achieving these goals depends on thoughtful and

Figure 2: CQI Cycle for SPP - Part I

[21] See the BERK Recommendations for Seattle's Preschool for All Initiative: http://murray.seattle.gov/wp-content/uploads/2014/05/BERK-Recommendations.pdf

[22] For a detailed description of the development of SPP, please see the Seattle Preschool Program Plan: http://www.seattle.gov/Documents/Departments/OFE/AboutTheLevy/EarlyLearning/SPP_ProgramPlan2015-16_Final.pdf

complete implementation of the SPP standards. DEEL, along with the Evaluation Team, will set targets and track indicators that measure progress toward achieving these standards.

The SPP Evaluation Strategy begins with identifying, reviewing, and operationalizing established program standards. These standards become the targets against which the evaluation measures progress. This section briefly describes the evidence base and precedent for SPP's standards and gives more detail about how the research questions are aligned to these standards.

SPP Program Standards

Table 3 describes key elements of the SPP program that inform the standard or the definition of quality.

Table 3: SPP standards for center and school-based providers

Program Element	Description of SPP Standards
Classroom Selection	• Classrooms will be selected from providers that meet program eligibility criteria. A critical aspect of provider eligibility will center on quality. All SPP providers must hold at least a Level 3 quality rating in Early Achievers (EA), Washington State's Preschool QRIS. • Priority will be given to agencies in areas where public elementary schools have records of low achievement, agencies that target services toward meeting the needs of preschool children from low-income families, and agencies that can provide evidence of high-quality practice or the availability of programs for DLL students.[23]
Classroom Quality and Improvement Expectations	• Because quality pre-K is the cornerstone of the Seattle Preschool Program, DEEL will ensure that providers improve quality over time. – After participating in SPP for two years providers are expected to be working toward an EA Level 4 rating. Teachers will be expected to have a CLASS combined Emotional Support and Classroom Organization rating that approaches 6.0 and an Instructional Support rating that approaches 4.5. For more information on CLASS, see Section III. – Where scores are not achieved and sufficient progress is not made, an SPP Coach will work with teachers to ensure that these goals are part of a Quality Improvement Plan (QIP).
Classroom Operating Schedule	• Preschool classes will operate on a full-day schedule: six hours per day, five days per week, and 180 days per year.
Class Sizes and Teacher Student Ratios	• Teacher-student ratio and class sizes will target an approximate ratio of one adult for every 10 students and a maximum of 20 students per classroom. • In the average classroom DEEL anticipates one lead teacher and one assistant teacher.
Curriculum	• Providers will adopt curricula approved by the city of Seattle, including HighScope or Creative Curriculum for Preschool, 5th Edition. • Teachers will be provided professional development and support for curriculum implementation and will be assessed on curriculum implementation fidelity.

[23] Head Start: http://eclkc.ohs.acf.hhs.gov/hslc/tta-system/cultural-linguistic/Dual%20Language%20Learners/DLL_%20Resources/OHSDefinitionof.htm.

Program Element	Description of SPP Standards
Dual Language Learners (DLL)	• The city of Seattle will support DLL classrooms by prioritizing DLL classrooms that are representative of the Seattle population. The program will strive to ensure a mix of children who are native speakers of each language in a classroom. • DEEL will aid instructional staff in achieving certification with bilingual endorsement and provide professional development on developmentally appropriate practices that support language acquisition. • DEEL will also ensure provider agencies assess children in the language(s) of instruction when feasible.[24]
Special Populations[25]	• Since the majority of preschoolers do not yet have a diagnosis of a behavioral or health need, SPP will conduct screening processes at least once per year. • Provider agencies may request additional funding (budgeted as "Special Populations Costs") through the DEEL coach assigned to the classroom when screenings or other information result in six or more children being designated members of a special population. • SPP provider agencies will adhere to all Individualized Education Plans (IEPs).
Other Services	• All SPP providers will be required to screen all children for developmental and behavioral concerns within 90 days of the start of the school year. • Providers must execute a culturally relevant plan for partnering with families and communities to improve child outcomes on an annual basis.
Instructional Staff Credentials and Compensation	• Lead teachers hired after the provider becomes an SPP participant will be required to have a bachelor's degree in early childhood education or a bachelor's degree and Washington State teaching certificate with a P-3 endorsement. • Other site staff have minimum education requirements as well, as specified in the Program Plan. Current staff will have four years to meet these requirements. Teachers will be offered advisory services and tuition support. • SPP is committed to high-quality support and a teacher compensation model that promotes retention and will work toward pay parity with salary guidance levels. Teachers will be paid according to their education and experience.

[24] A limitation of the assessments is that they are normed in Spanish and English.
[25] Seattle Preschool Action Plan (p. 20).

Program Element	Description of SPP Standards
Teacher Professional Development (PD)	• Teacher professional development will be fully funded by SPP, aligned with other preschool programs that DEEL manages, such as Head Start, ECEAP and Step Ahead, and will leverage WA DEL trainings. • All staff will receive preservice training provided by DEEL coaches. – Trainings will focus on program standards, processes, and principals; screenings and assessments; and creating and sustaining safe and supportive preschool learning environments. – Additional content training will be delivered throughout the year by SPP coaches. – Coaches will also provide instructional support for the following program areas: teaching children whose primary home language is not English; teaching children with special needs including behavioral needs; teaching in a culturally and linguistically responsive manner; and partnering with families. – Professional development will also be available on strategies, policies and practices that are consistent with the city of Seattle's RSJI principles, creating and sustaining a language-rich classroom environment, and best practices in reflective coaching, educational leadership, and business management for directors and program supervisors. • Additional trainings will be provided as needed.
Coaching Model	• The SPP coaching model will target one coach for every 10 classrooms for the demonstration phase, which is between 2015 and 2018. Coaching hours will be allocated based on assessments, observations, and QIPs. • Coaches will provide teachers support for critical program elements, including supporting the implementation of curriculum with fidelity, using assessments and teacher observation to strengthen teacher practice, helping teachers individualize instruction, working with teachers to integrate emergent approaches for curriculum and children's interests, using child and classroom assessments and PD plans to inform practice, and providing support in working with children with special needs, including behavioral health needs.

Evaluation research questions by SPP program standards

The DEEL Team and the external Evaluation Team will work together to measure SPP's progress in achieving the program standards. To support this, the Evaluation Team constructed a set of research questions that will focus the evaluation on the elements of quality that need to be measured and analyzed to ensure that SPP is meeting the program standards. DEEL and the Evaluation Team will define the associated targets for each of these questions over the course of the first year.

See Table 4 for a list of the research questions and the evaluation methods that will be used to answer, measure, or validate each question.

Table 4: SPP standards and evaluation strategy research questions

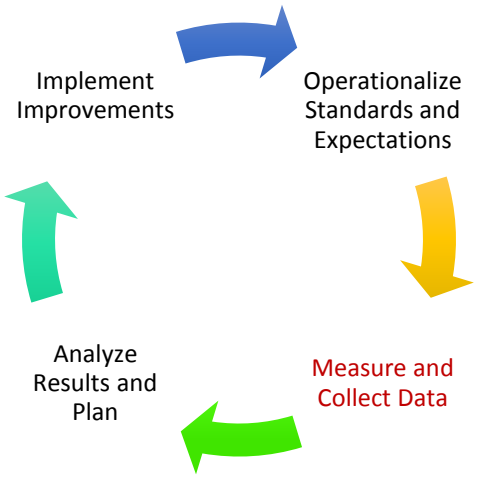
Program Element	Questions to Determine if Providers are Meeting SPP Standards	Evaluation Method
Classroom Selection	• What are the demographic characteristics of the children enrolled in SPP? How do they compare to the demographics of children in Seattle more generally?	Impact Evaluation
	• Is the SPP serving students from targeted communities?	Self-Evaluation
Classroom Quality and Improvement Expectations	• What is the overall observed quality of children's interactions with teachers, each other, and the physical environment each year? (Quality includes keeping children safe as well as attending to their emotional, social, and cognitive needs.) • What activities do children engage in, and is there scope for their interests and active participation? • How much does the quality of children's experiences provided by SPP improve from year to year? • How do these vary within SPP across children and providers?	Impact Evaluation
	• Instructional Quality: What is the quality of teacher-student interactions? • Environmental Quality: Is the classroom conducive to student learning, and does it contain the tools and supplies for teacher effectiveness?	Process Evaluation
	• Are sites meeting quality thresholds, improving their quality, and engaging with DEEL coaches to understand and improve quality?	Self-Evaluation
Classroom Operating Schedule	• Are sites operating programming for six hours per day, 180 days per year?	Self-Evaluation
Class Sizes and Teacher Student Ratios	• What were the reported levels of child attendance during the pre-K year for SPP children each year?	Impact Evaluation
	• Are sites sustaining appropriate class sizes and teacher/student ratios?	Self-Evaluation
Curriculum	• Is the program using evidence-based curriculum? Is the program sticking to the curriculum (fidelity of implementation)? • Are teaching practices aligned with learning standards?	Process Evaluation
	• Have sites purchased and implemented approved curriculum with fidelity?	Self-Evaluation
Instructional Staff Credentials and Compensation	• Are teachers qualified based on SPP standards or working toward qualifications? • Are teachers compensated in keeping with program standards?	Process Evaluation
	• What are teacher qualified levels? • What are teacher compensation levels?	Self-Evaluation

Teacher Professional Development (PD)	• Do teachers have the appropriate skills and knowledge needed to provide high-quality instruction? • Is professional development aligned with SPP design standards, processes, and principles?	Process Evaluation
	• Are administrative and instructional staff completing appropriate professional development according to standards (i.e., curriculum, etc.)?	Self-Evaluation
Coaching Model	• Are sites engaging appropriately with coaches and DEEL staff in continuous improvement?	Self-Evaluation
DLL	• Are sites providing supports for DLL students?	Self-Evaluation
Special Populations	• Are sites providing services for students from special populations?	Self-Evaluation
Other Services	• Are sites providing support services for students?	Self-Evaluation

Section III: Measure and Collect Data

This section describes the data and measurements the Evaluation Team will collect. Prior to outlining the measurement and data collection plan, this section provides a brief review of the literature that guided measurement planning.[26]

Figure 3: CQI Cycle for SPP - Part II



Lessons Learned on Measurement

There is substantial, peer-reviewed research on best practices in collecting data for an impact evaluation to assess the effects of early learning programs on children.[27] To augment the limited literature on process evaluation and self-evaluation of high-quality public preschool programs, the Evaluation Team conducted interviews with representatives from two exemplar preschool programs, Boston and New Jersey, to identify best practices for these two methods.[28]

Impact evaluation

Evaluations of high-quality public preschool programs have been similar with respect to the information they have collected on children, families, and other program standards. Successful evaluations assess

[26] Note that this section does not address sampling and research design for the impact evaluation. Section IV: Analyze Results and Plan provides detail on sampling and research design.
[27] Nores, M., et al. (2015) Technical Report for the City of Seattle. A Review of the Evidence on Preschool Programs and a Comparison of Selected State and City Programs.
[28] CEELO: http://ceelo.org/wp-content/uploads/2013/12/State-Pre-K-Monitoring-and-Evaluation-Policies.pdf.

children using standardized measures of language, literacy, math, and, more rarely, executive functions (e.g., self-control and attention), social skills, and behavior. "Standardized" means that the measures were developed to ensure they are valid, reliable, accurate, fair, and used consistently regardless of who collects the information. Successful evaluations also collected detailed information on family background, especially as it relates to home learning and preschool participation. These programs collect this information from families by October.

Although the assessments typically used in preschool evaluations are not available in languages other than English and Spanish, they have been shown to work well in measuring impact on children from diverse backgrounds.[29] In addition, collecting information on family perspectives about how the program affects their children increases the reliability of the impact evaluation for families from diverse cultural and language backgrounds. Insights from families allow the Evaluation Team to make culturally relevant analyses and interpretations of findings.

The most effective evaluations use multiple assessments from multiple perspectives to measure the impact of preschool programs on children. The research indicates that impact evaluations should assess, at a minimum: language, literacy, math, executive functions, and social skills and behavior.

Process evaluation

Process evaluation typically measures the program standards that support effective teaching because observed classroom quality is correlated with improvements in children's learning and development.[30] Classroom quality is multidimensional, but high-quality student experiences are at its core. These experiences can be measured by observing the quality of teaching, how the teacher interacts with children and guides their activities, and how well the classroom is set up for quality teacher-student interaction and learning.[31] Direct observations are widely used in preschool evaluation studies, child care quality rating systems, and by program administrators. Effective teacher-student interactions, such as those measured by the CLASS and ECERS-3, provide critical input on the quality of the child's experience through measures of emotional support, classroom organization, instructional support, and environmental supports.[32]

Although most evaluations have used only a single observation measure to assess classroom quality,[33] the research indicates that evaluators should collect multiple observation measures of the classroom. At least one of these measures should focus on the experiences of an individual child.[34] Other common

[29] Assessments are not suitable for children with severe disabilities.
[30] Reynolds, Arthur J., Temple, Judy A., Ou, Suh-Ruu, et al. (2007). Effects of a School-Based Early Childhood Intervention on Adult Health and Well-Being: A 19-Year Follow-Up of Low-Income Families, *Archives of Pediatric and Adolescent Medicine*, *161*(8), 730–739.
[31] Zaslow, M., Martinez-Beck, I., Tout, K., and Halle, T., (2011). Quality Measurement in Early Childhood Settings, Baltimore, MD: Brookes Publishing.
[32] Measuring and Improving Teacher-Student Interactions in PK-12 Settings to Enhance Children's Learning; University of Virginia CASTL; www.curry.virginia.edu/casl.
[33] Appendix A.
[34] This would be the case for example of the Snapshot (CITE).

measures to improve quality, such as class size, teacher education and credentials, and curriculum, are critical, but evaluation must also assess the academic and social environment.[35]

According to the researchers who evaluated high-quality public preschool programs, an evaluation of a nascent preschool program should begin with a targeted process evaluation.[36] While it takes time to accumulate data for the process evaluation, it is important to begin measuring immediately to develop a culture of using data for continuous improvement. Both Boston and New Jersey, two public programs, used results from their initial process evaluations to make course corrections to their programs. Boston used results to implement a standardized curriculum in the classroom. New Jersey honed its professional development program to provide support for teachers in areas it found to need improvement.

Self-evaluation

Preschool programs are more likely to have consistently good teaching if they have qualified teachers who receive ongoing support, supervision, and regular feedback from experts on how they are teaching. Some publically funded preschool programs have used self-evaluations to collect information that can be used to identify areas for improvement for teachers. The self-evaluation is not used to judge individual children or teachers but as a basis for feedback to constantly improve children's experiences in preschool.[37]

New Jersey, a well-documented, high-quality program, successfully implemented a self-evaluation tool called the Self-Assessment Validation System (SAVS). The SAVS is an objective rubric that providers used to measure key elements of program implementation and teacher-child interaction. New Jersey, which did not have a QRIS at the time, developed SAVS in collaboration with its provider community, and coaches and providers worked together to use the tool to assess providers' performance.[38] As a result, the tool had more credibility with providers, and they were more likely to trust and use the results to implement improvements.[39] These benefits are especially important in a system that is adapting multiple programs and employing professionals with a wide range of skills and experience.

The SPP Evaluation's Approach to Measurement

DEEL and the evaluation will include impact, process, and self-evaluation indicators at the program, classroom, and child level. Evaluation data and measurements will come from a range of sources, including primary sources (classroom and child assessments administered by the Evaluation Team, TS GOLD data, surveys, and interactions between DEEL coaches and education specialists, teachers, directors, students, and families) and secondary sources (DEEL and WA DEL enrollment and teacher professional development datasets). The Evaluation Team will collect quantitative and qualitative data

[35] Improving Teacher-Student Interactions in PK-12 Settings to Enhance Children's Learning.
[36] Interview with Jim Minervino, 6/1/2015.
[37] Technical Report for the City of Seattle. A Review of the Evidence on Preschool Programs and a Comparison of Selected State and City Programs.
[38] Interview with Ellen Frede, PhD, 6/17/2015.
[39] Ibid.

because neither one is sufficient by itself. Quantitative data provides fodder for analyses of trends and improvement, and qualitative data allows the Evaluation Team to talk about why those trends occur.

When data come from existing data sources, such as WA DEL, the Evaluation Team will ensure the data are reliable, valid, and objective before using it.

Members of the Evaluation Team who have extensive experience with and training in surveys, focus groups, and other methods of qualitative data collection will administer surveys, interviews, and focus groups.

Conducting culturally competent assessments

Conducting culturally competent assessments requires that assessment personnel integrate culturally competent attitudes, knowledge, interview skills, intervention strategies, and evaluation practices specifically informed by the age, culture, and language of the child. The Evaluation Team has expertise in leading assessments in publically funded preschool programs with diverse populations.

Assessment personnel will be trained on the relevant assessment and have prior experience working with children, knowledge of standardized assessment, and education in early childhood development or related fields.

Where possible, assessment personnel will be selected to match the culture, race, and languages of children being assessed. While child assessments are only available and normed in Spanish and English, assessment personnel will provide verbal instructions in the student's home language using an interpreter, if necessary.

Assessments described in this section have been validated for use with children whose ages, cultures, socioeconomic status, abilities and disabilities, and other characteristics are similar to the children who will be assessed for the SPP evaluation. Given that child assessment tools have not yet been developed and normed in languages other than Spanish and English and the cost for translating and norming these assessment tools is beyond the scope and capacity of the city of Seattle and its evaluators, the Evaluation Team will use tools that have been used in other evaluations of publically funded, racially, culturally, and linguistically diverse preschool programs.

Procedures relating to data collection reliability, validity, and security

The University of Washington and NIEER will oversee student assessments. The following bulleted list summarizes the main premises under which assessment will be carried out. The Evaluation Team and DEEL will follow these requirements to ensure data security and validity and safety for children and families:

- The measures chosen have been used in a variety of cultural contexts and countries and have been effective at capturing child growth and development overall and for particular programs.
- Experts will help develop assessment protocols and will change or modify them if necessary, based on feedback from directors, teachers, or families.
- All assessment personnel will have Human Subjects Certification and have passed background checks.

- The assessment personnel will be trained on timeliness, culturally competent behavior and etiquette, and gauging child discomfort and level of engagement. They will also be provided with letters of introduction for directors and teachers and be provided resources that support their knowledge and skills about culturally and linguistically appropriate early childhood assessment.
- To the extent possible, the Evaluation Team will match the cultural and language composition of assessment personnel to the composition of children to be assessed.
- For language, literacy, math, and social-emotional assessments, the Evaluation Team will use only existing normed assessments, currently available in English and Spanish. For assessments of executive function, the Evaluation Team will use existing assessments for which an adaptation is not necessary beyond having the instructions be delivered in the child's primary language.
- All assessments will be done one-on-one with children. The assessment personnel will record responses on password-protected tablet computers. Information from the tablet computers will be shared with other members of the Assessment Team through encrypted emails. All data will be kept strictly confidential. Data are safeguarded from either accidental or intentional access from unauthorized persons.
- The assessment personnel will distribute informed consent/assent forms to all parents in different languages, as needed. Staff will be available to explain the form in multiple languages. Informed consents include information on the study and its goals, what participation implies for the families, risks and benefits, duration, incentives, if these are part of the study, freedom to withdraw, explicit assurance of participant's confidentiality/anonymity in investigator's reports of findings, and information on contact persons. Consent forms require Institutional Review Board (IRB) approval before they are used in the field. The Field Collection Team will partner with SPP in efforts to communicate to parents.
- Information on the child's home environment is collected through family surveys. These allow appropriate and relevant interpretation of the information across children of different linguistic and cultural backgrounds.
- Participant identities will be kept strictly confidential in all published or publicly available documents. Children will be identified by a secure code rather than their names or any other identifying information throughout the course of the project.

The assessment personnel will assess children individually, paying particular attention to the child's level of engagement and providing breaks as needed. The complete assessment is expected to take 20–25 minutes per child. To help ensure the reliability and validity of assessments:

- Child assessments will be conducted one-on-one and are to be scheduled to avoid meal, nap, and outdoor play times.
- To increase response rates to family surveys, the Evaluation Team will distribute incentives to families, such as gift cards to local stores.

Institutional Review Board

Once approved by the Seattle City Council, the Evaluation Team will submit a study protocol to the Rutgers University Institutional Review Board (IRB) for ethics approvals. IRB approval requires a clear

and well-developed evaluation plan. The entire IRB approval includes developing consent forms, outlining the plans for data collection, analysis, and dissemination, and other requirements. In addition, the IRB requires submission of yearly continuation requests, as well as any changes to the data collection protocol (process), instruments used, sampling frame, and use of incentives.

Using third-party data

The SPP Evaluation Strategy leverages WA DEL and other partner data where possible. This requires gathering, cleaning, and maintaining existing data along with new data sources. The Evaluation Team has experience extracting, cleaning, loading, and analyzing data from WA DEL data systems, including teacher credentials and professional development data from MERIT and detailed provider ratings data from PRISM.

For WA DEL and DEEL data, the Evaluation Team will perform a data discovery process to ensure the quality of the data from all third-party datasets. Any issues with data quality will be referred to WA DEL and DEEL to mitigate and resolve in a timely manner.

All data and measures will be consolidated into a relational data management system in order to support analysis and reporting and will be provided to DEEL to support internal analysis and reporting needs.

Measurement Tools and Techniques

The measurement tools and techniques can be separated into three broad categories: observations of classroom quality, child assessments, and qualitative data collection tools. Table 5 summarizes each data collection tool, the evaluation method it supports, who will use the tool, and anticipated timing for using the tool.

Table 5: SPP evaluation strategy measurement summary

Category	Tools/Assessments	Data Collection Evaluation Method	Data Collection Team	Data Collection Anticipated Timing
Classroom Quality and Teacher-Student Interactions	CLASS	Impact evaluation	Evaluation Team	Spring
		Self-evaluation	DEEL coaches	Fall
	ECERS-3	Impact evaluation	Evaluation Team	Spring
		Self-evaluation	DEEL coaches	Fall
	Site-level assessment rubric	Self-evaluation	Site administrators	Ongoing
	WA DEL Managed Education and Registry Information Tool (MERIT)	Process evaluation	Evaluation Team	Monthly
	PRISM	Process evaluation	Evaluation Team	Monthly

Category	Tools/Assessments	Data Collection Evaluation Method	Data Collection Team	Data Collection Anticipated Timing
Child Assessments	Peabody Picture Vocabulary Test IV (PPVT IV) / Test de Vocabulario e Imágenes Peabody	Impact evaluation	Evaluation Team	Fall and Spring
	Woodcock-Johnson Tests of Achievement, 3rd Edition/ Bateria Psico-Educativa Revisada de Woodcock-Muñoz (WM-R)	Impact evaluation	Evaluation Team	Fall and Spring
	Peg Tapping Task; Dimensional Change Card Sort; Head-Toes-Knees-Shoulders; Task Orientation Questionnaire; Child Behavior Checklist (CBCL)	Impact evaluation	Evaluation Team	Fall and Spring
	TS GOLD[40]	Impact evaluation	Evaluation Team	Fall, Winter, and Spring
		Self-evaluation	DEEL coaches	Fall, Winter, and Spring
Qualitative Data Collection Tools	Center director surveys	Process evaluation	Evaluation Team (with DEEL)	Winter
	Teacher surveys	Process evaluation	Evaluation Team	Winter
	Coach focus groups	Process evaluation	Evaluation Team	Winter
	Participating families surveys	Impact evaluation	Evaluation Team	Fall

The next section describes the measurement tools and techniques in greater detail, including the specific data that will be collected. DEEL and the Evaluation Team will develop a work plan each summer that more clearly determines how much data collection is done by the coaches and the Evaluation Team in order to manage workload.

Details on SPP measurement tools, techniques, and uses

Table 6 describes each measurement tool and who will collect the data. Subsequent sections provide more detail about each assessment in Table 6, including how it will be used to support the SPP evaluation.

[40] TS GOLD is administered by teachers. In this case, "collecting data" refers to gathering data entered by teachers into a TS GOLD database.

Table 6: SPP evaluation strategy measurement detail

Category	Tools/Assessments	Brief Description of Tool/Assessment	Brief Description of Data Collection Personnel
Classroom Quality and Teacher-Student Interactions	CLASS	CLASS measures the interactions between students and teachers in areas that are linked to student achievement and development.	Assessment personnel (impact evaluation) and DEEL coaches (self-evaluation) will be trained to obtain valid and reliable information from children and classrooms. Assessment personnel are trained to minimize disruptions to the classroom when assessing individual children and during observations. They blend into the background and avoid unnecessarily interacting with children or teachers or interfering with activities.
	ECERS-3	ECERS-3 is used to assess the quality of the classroom environment.	
	Site-level assessment rubric	An objective rubric is used to collect provider-level data and for providers to rate or score their own performance.	DEEL coaches and education specialists work with providers to collect data. Site administrators and directors will be the primary data source.
	WA DEL Managed Education and Registry Information Tool (MERIT)	MERIT contains real-time teacher-level data on: • Teacher qualifications and credentialing • Teacher educational attainment data • Professional development hours by training area	The Evaluation Team will collect the data from WA DEL.
	PRISM	The PRISM data warehouse contains EA ratings data by provider, including detailed scores on individual items assessed in a rating.	The Evaluation Team will collect the data from WELS.
Child Assessments	Peabody Picture Vocabulary Test IV (PPVT IV)/ Test de Vocabulario e Imágenes Peabody	The PPVT IV is a broad assessment of what the child understands by measuring receptive vocabulary.	Trained assessment personnel will administer child assessments.

Category	Tools/Assessments	Brief Description of Tool/Assessment	Brief Description of Data Collection Personnel
	Woodcock-Johnson Tests of Achievement, 3rd Edition/ Bateria Psico-Educativa Revisada de Woodcock-Muñoz (WM-R)	The Woodcock-Johnson Test assesses children's mathematical and literacy skills. In particular, these assessments focus on letter and word knowledge and applied math.	
	Peg Tapping Task; Dimensional Change Card Sort; Head-Toes-Knees-Shoulders Task; Task Orientation Questionnaire; Child Behavior Checklist (CBCL)	Measures of cognitive, socio-emotional outcomes, and executive function.	
	TS GOLD	TS GOLD measures developemental readiness on six fundamental ares and is aligned with the Common Core State Standards, state early learning guidelines, and the Head Start Child Development and Early Learning Framework.	Teachers perform the TS GOLD assessment and enter data into a TS GOLD database. DEEL will have access to the database and share data with the Evaluation Team at the discretion of the team.
Qualitative Data Collection Tools	Center director surveys	Survey or focus group with center directors to understand their experiences and perspectives on their center's implementation of the SPP.	The Evaluation Team will work closely with DEEL to collect this information.
	Teacher surveys	The teacher survey will gather teachers' perspectives on their experiences with the SPP.	The Evaluation Team will collect the data from teachers directly.
	Coach focus groups	The Evaluation Team will interview coaches to understand their perspectives on the implementation of the SPP with providers.	The Evaluation Team will facilitate the focus groups.

Category	Tools/Assessments	Brief Description of Tool/Assessment	Brief Description of Data Collection Personnel
	Participating families surveys	Data from the families will be collected using an existing questionnaire used for the WA DEL EA program.	Directors will contact families in their preschools and invite them to participate. The Evaluation Team will collect the data; providers will never see the individual survey responses.[41]

Classroom Assessment Scoring System (CLASS)

CLASS is used to assess classroom practices in preschool through third grade by measuring the interactions between students and adults. These practices are broadly grouped into three domains: instructional support, social/emotional climate, and classroom management. These can be further subdivided into finer-grained measurements of the quality of interactions. For example, the emotional support domain is composed of: positive climate, negative climate, teacher sensitivity, and regard for student perspectives.

CLASS uses a seven-point scale on which a score of 1 or 2 indicates low-range quality and a score of 6 or 7 indicates high quality. Each dimension and domain is assigned a score during each 20-minute observation period (observers watch for 20 minutes and then record for 20 minutes in cycles). The number of children and adults in the classroom is also recorded during each 20-minute cycle. Appendix C provides descriptions of each CLASS dimension.

CLASS is widely used in pre-K classrooms because it describes multiple dimensions of teaching that are linked to student achievement and development and has been validated in more than 2,000 classrooms. Studies demonstrate the reliability of CLASS scores across observers, cycles, days, and school years.[42] CLASS and ECERS-3 predict each other to some extent but not so highly that they can be considered to measure the same things. Within Washington, CLASS is already used to assess quality in the state's EA program.

The CLASS assessment will provide the Evaluation Team with an impartial measure of instructional quality. The Evaluation Team will collect CLASS data independently from DEEL coaches, although it will use the assessment data for similar purposes. Additionally, feedback from the CLASS assessment will be used in the self-evaluation to provide DEEL coaches with a baseline for classroom quality and a starting point and plan for where to focus coaching effort to improve quality.

Early Childhood Environment Rating Scale - Third Edition (ECERS-3)

As with the CLASS measurement, the ECERS-3 measures the quality of the classroom environment on a seven-point scale, indicating a range of quality from inadequate (1) to excellent (7). The ECERS-3 is

[41] The city will ensure that all families have the opportunity to participate. In instances where parents or guardians are illiterate, the survey will be given as an interview.

[42] Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher*, 38, 109-119.

composed of 35 items grouped into six subscales: Space and Furnishings, Personal Care Routines, Language-Reasoning, Learning Activities, Interaction, and Program Structure. ECERS-3 is the newest observation tool in the field, but its predecessor the ECERS-R has been widely used and has well-established validity and reliability. The previous version, ECERS-R, is currently used to assess quality in the state's EA program. At the appropriate time ECERS-3 will be implemented in Washington State. The ECERS-3 provides more insights into the content of what is taught in preschool—rather than simply how it is taught—than previous versions.

Each subscale is assigned a score during each 20-minute observation period (observers watch for 20 minutes and then record for 20 minutes in cycles). Appendix C provides descriptions of each ECERS-3 dimension.

The ECERS-3 assessment will provide the Evaluation Team with an impartial measure of the instructional environment. The self-assessment will use the data from the ECERS-3 the same way it will use data from CLASS, to target coaching to the needs of individual teachers.

Site-level assessment rubric

DEEL and providers will develop an objective rubric to collect data about the implementation of the SPP Standards. In particular, the rubric will be used to assess the extent to which providers are implementing program standards related to:

- Curriculum implementation
- Curriculum training
- Integration of Early Learning Standards
- Use of data to make improvements
- Class size
- Teacher-student ratio
- Standards
- Program hours
- Meeting the needs of children in Special Populations, as defined in the SPP Implementation Plan
- Meeting the needs of DLL students
- Cultural relevancy of the classroom's activities and environment

Data from the rubric will give the DEEL Coaching Team continuous feedback on areas in which providers have opportunities to improve. Coaches can then adjust their coaching and technical assistance strategies to meet provider needs.

WA DEL Managed Education and Registry Information Tool (MERIT)

MERIT data contain teacher-level records with information about professional development hours, educational attainment (degrees and credentials), teacher demographic data, and employment information. Data from MERIT will be used for two separate purposes.

- Track teacher credentials: The Evaluation Team will use MERIT data to track progress toward degree and credential requirements each year. MERIT will include information about degrees and credentials for newly hired teachers as well.
- Track professional development hours: The Evaluation Team will measure all professional development hours entered into MERIT. When available, the Evaluation Team will also use MERIT data about the specific content of the training.

The Evaluation Team will rely on SPP teachers registering in MERIT (as required by the program) and appropriate support from WA DEL to maintain, clean, and share MERIT data with the Evaluation Team.

DEEL will partner with the WA DEL to track teachers' educational attainment, advisory services, and tuition reimbursment through MERIT. DEEL will use this information in the self-evaluation to measure progess over time toward standards related to teacher qualifications, assess barriers to progress, and measure the effectiveness of incentives.

PRISM

The PRISM data warehouse tracks data for all the standards associated with EA. It is built from the WELS system. WELS was designed for early childhood school administration agencies to track the quality of child care centers.

The Evaluation Team will collect EA ratings for SPP sites, including CLASS and ECERS-R[43] data from PRISM. This will provide a baseline to measure improvement over time as SPP classrooms fully implement program standards.

PRISM includes item-level scores, which allows evaluators to observe specific elements of quality and analyze patterns in quality within and across providers. PRISM data include:

- Detailed CLASS rating details by classroom for each provider
- Detailed ECERS-R rating details by classroom for each provider

EA ratings are conducted every three years but not on every classroom. The Evaluation Team will only use recent and pertinent data in this evaluation.

Peabody Picture Vocabulary Test IV (PPVT-IV)/ Test de Vocabulario e Imágenes Peabody

The Peabody Picture Vocabulary Test IV and its Spanish counterpart, the Test de Vocabulario e Imágenes Peabody, assess language development. The PPVT-IV, which measures receptive vocabulary, is considered a broad assessment of what the child understands in his or her language. Assessment personnel present a set of four images, from which the child picks the image they think represents the word they hear. The measure is considered valid and reliable.[44] The PPVT-IV is the most frequently

[43] ECERS-R is currently used by Washington State EA; ECERS-3 replaces ECERS-R.

[44] Reliability refers to the degree to which a measure is consistent. A measure is said to have a high reliability if it produces similar results under consistent conditions. Validity refers to the accuracy of an assessment—whether or not it measures what it is supposed to measure.

administered literacy assessment; it has been normed in both Spanish and English on large numbers of children.

Woodcock-Johnson Tests of Achievement, 3rd Edition/ Bateria Psico-Educativa Revisada de Woodcock-Muñoz (WM-R)

The Woodcock-Johnson Tests of Achievement, 3rd Edition, and the Bateria Psico-Educativa Revisada de Woodcock-Muñoz (WM-R) will assess children's mathematical and literacy skills development. These measures, which have been used in numerous large-scale preschool studies, are consistently reliable and valid.[45]

Peg Tapping Task; Dimensional Change Card Sort; Head-Toes-Knees-Shoulders (HTKS); Task Orientation Questionnaire

Assessment personnel will use at least three of the following tasks to measure executive function:

- Peg Tapping Task: Measures cognitive inhibitory control. It has shown high predictive validity on the Vanderbilt study and great performance.[46] It's available in English, and NIEER has a Spanish translation that it has used in the field for at least two years.
- Dimensional Change Card Sort: Measures attention shifting. It's available in English, and NIEER has a Spanish translation that it has used in the field for at least two years.
- Head-Toes-Knees-Shoulders (HTKS) Task: Measures inhibitory control and attention. The HTKS task has been widely used; it is also a consistent predictor of emergent mathematics, vocabulary, and literacy in preschool children. It's available in English, and NIEER has a Spanish translation that it has used in the field for at least two years.
- Task Orientation Questionnaire: Measures compliance and attention. The scale shows predictive validity of cognitive and socio-emotional outcomes and executive function measures and has been validated in the U.S. The questionnaire is also available in Spanish.[47]

The Child Behavior Checklist has a teacher- and a parent-rated, low-cost form that could be used to measure children's emotional, social, and behavioral development. Available in English and Spanish, this measure would serve as a complement to the proposed battery.[48]

---

[45] Early, D. M., Maxwell, K. L., Burchinal, M., Alva, S., Bender, R. H., Bryant, D., Cai, Karen, Clifford, R., Ebanks, C., Griffin, J., Henry G., Howes, C., Iriondo-Perez, J., Jeon, H., Mashburn, A., Peisner-Feinberg, E., Pianta, R., Vandergrift, N., & Zill, N. (2007). Teachers' education, classroom quality, and young children's academic skills: Results from seven studies of preschool programs. *Child Development*, 78(2), 558–580.

[46] Lipsey, M., Nesbitt, K., Farran, D., Dong, N., Fuhs, M., & Wilson, S. (2014, May 1). Learning-Related Cognitive Self-Regulation Measures for Prekindergarten Children with Predictive Validity for Academic Achievement (Working Paper). Retrieved June 10, 2015, from https://my.vanderbilt.edu/cogselfregulation/files/2012/11/Self-Reg-summary-paper-5-7-141.pdf.

[47] Smith-Donald, R., Raver, C. C., Hayes, T., & Richardson, B. (2007). Preliminary construct and concurrent validity of the Preschool Self-regulation Assessment (PSRA) for field-based research. *Early Childhood Research Quarterly, 22*(2), 173–187.

[48] Achenbach System for Empirically Based Assessment: http://www.aseba.org/preschool.html

For more information on student assessments that may be used in the classroom, although they are not necessarily involved in the demonstration phase evaluation, please see Appendix D.

Student TS GOLD scores

TS GOLD measures developmental readiness on six fundamental areas of development: social-emotional, physical, language, cognitive, literacy, and math. Teachers use the results of TS GOLD to tailor instruction. This assessment, unlike others described above, is administed by the teacher rather than the Seattle Evaluation Team. TS GOLD results will demonstrate whether or not teachers are using student assessment data to improve and customize instruction.

TS GOLD is used by teachers in many classrooms in Seattle, including ECEAP and some Head Start programs. For more information on the areas of development and learning assessed by TS GOLD, please see Appendix C.

Center director surveys

The Evaluation Team will conduct surveys with center directors in 2016–2017. Given the low volume of participating centers in Year 1, the Evaluation Team will do in-depth interviews with select directors to capture findings in Year 1 and inform the survey protocol.

Center director surveys will be used to measure and understand the role of center directors in the implementation of SPP program standards, with a specific focus on those standards that impact teacher-student interactions. Surveys are intended to supplement and provide input on the internal evaluation work described above. The surveys will ask questions such as: How is the center director spending time and resources? To what extent do DEEL program policies and requirements support program implementation and improvement? Is DEEL providing center directors with the resources necessary to successfully implement the SPP?

Teacher surveys/interviews

The Evaluation Team will conduct teacher surveys or interviews in 2016–2017. Given the low volume of centers in Year 1, the Evaluation Team will conduct in-depth interviews with select teachers to capture findings in Year 1, inform the decision of whether surveys or interviews are appropriate, and to inform the content of the protocol.

Teacher surveys will be used to measure and understand teachers' implementation and integration of SPP program standards. They will also measure how program standards work together to improve student learning. Surveys are intended to supplement and provide input on the internal evaluation work described above. The surveys will ask questions such as: How are teachers using data to inform instruction? To what extent is DEEL providing teachers with technical assistance and other supports necessary for the teacher to be successful?

Coach focus groups

Coach focus groups will be used to measure and understand the coach's role in supporting implementation of SPP program standards at SPP providers. In the first year of the SPP, the Evaluation Team will hold informal discussions about progress, provider technical assistance needs, and DEEL capabilities and resources. Beginning in 2016–2017, as the program scale expands, the Evaluation Team

will facilitate more formal focus groups with DEEL coaches. In the first year of the SPP the Evaluation Team will partner closely with DEEL to create the moderator guide for focus groups.

Data from the coach focus groups will be used to measure how effective coaching and technical assistance are in helping providers incorporate SPP program standards and improve quality.

Participating families survey

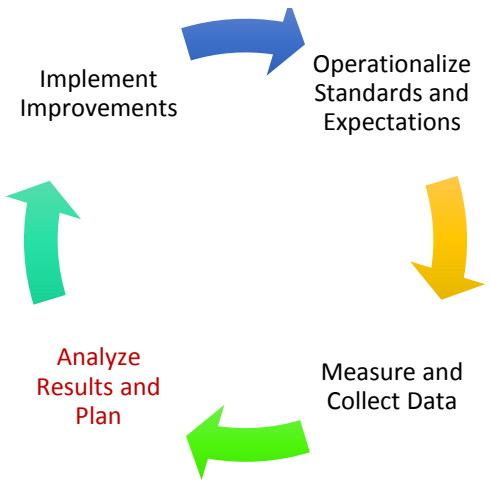
To gather information about the child's out-of-school environment, the Evaluation Team will develop and administer a family survey. The instrument will be based on a current family survey used by the University of Washington's Center for Childcare Quality and Early Learning as part of Early Achievers in Washington State. The survey will be optional, respecting a family's right to privacy. All survey responses will be kept confidential and used only by the Evaluation Team to help increase classroom and program quality. Families will be asked to participate in order to support program quality and their child's experience.

- Basic demograhics of the child and family such as family income, education, employment status, marital status, race/ethnicity, languages spoken at home, family structure, and family size.
- Learning activities in the home and other types of care and education the child may receive outside the home.
- Family perceptions of early education or child care programs, school attendance, and family perspectives on the benefiits of SPP including impacts on their child's learning and development.

In general, the family questionnaire (or interview, when needed) will use questions that have been used in representative national studies, allowing valid comparisons to comparable families in national datasets.

Section IV: Analyze Results and Plan

Figure 4: CQI Cycle for SPP - Part III



Analyze results and plan is the point in the CQI cycle when evaluators use statistics and other methods to generate results that can be used to inform decision making. For the SPP Evaluation, results from the analyses will help DEEL and providers identify opportunities for improvement and allow DEEL, the Seattle Mayor's Office, and the Seattle City Council to track SPP's progress toward achieving its goals. DEEL and the Evaluation Team will ensure that analyses are utilization-focused, or intended to be actively used by the Program Team and other stakeholders.

SPP Evaluation's Approach to Analyzing Results

Impact evaluation

The purpose of the SPP impact analysis is to determine the changes in student outcomes that can be attributed to SPP. The Evaluation Team will assess growth for children enrolled in SPP, compare this growth across years, and compare children's gains in the program to gains made by similar children who do not attend. This section describes the SPP impact evaluation analysis and alternatives for comparing learning and development for children enrolled in SPP to children not enrolled.

Study design

The Evaluation Team will use a combination of three components to conduct the impact analysis:

- Pre-Post Design: Assess growth for children in SPP by measuring learning development at the beginning and at the end of the year. This is called a pre-post comparison.
- Overlayed Comparative (RCT) Design: Compare growth between children who do and children who do not attend SPP. In accordance with SPP enrollment priorities, DEEL will determine which applicants will attend SPP and which are wait-listed.
- Relate quality of the classroom environment to children's pre-post gains using the self-evaluation and process evaluation measures of classroom.

The combination of these three approaches will permit the Evaluation Team to assess growth for children in the program, compare this growth across years and as programs mature, and compare children's gains in the program to gains for similar children who do not attend. Each of these designs are outlined in more detail below.

Pre-post design

Using student outcomes at the beginning and at the end of each preschool year is the simplest and easiest way to measure program performance. The major limitation of this approach is that all young children will progress over that period of time—even those who do not attend preschool. Gains in student outcomes are caused by many things, including combined effects of the child's biological development, home, and neighborhood, as well as attending SPP. The Evaluation Team will isolate the gains associated with SPP by comparing growth across years and across classrooms that have participated in SPP for varying lengths of time. In addition, the Evaluation Team will use measures that have been used in evaluations of other preschool programs so that gains in SPP can be compared to gains elsewhere.

The Evaluation Team will aim for a sample size of as many as 400 children and families. As SPP adds classrooms each year, the Evaluation Team will sample fewer children from each classroom while increasing the total number of children. This design will allow the Evaluation Team to compare progress in child care centers with two years in SPP to progress of children in centers with one year in the SPP, maintaining a representative sample of the system. The information collected will allow the Evaluation Team also to anchor any formative assessments used within the centers, such as TS GOLD.

Table 7 depicts the impact evaluation design. Each row in this table represents a different cohort of centers/classrooms being integrated into SPP. Each column reflects a new academic year. The cells

represent the number of children in each year and from each cohort. In the first year (2015–16), the Evaluation Team will collect data only for the first set of classrooms early in the fall (pre-test) and then again in late the spring (post-test).

Table 7: Pre-post design

	2015–2016	2016–2017	2017–2018	2018–2019
	All students	*8 per classroom*	*4 per classroom*	*4 per classroom*
SPP Classrooms Cohort 1	280 students	112 students	56 students	56 students
SPP Classrooms Cohort 2		200 students	100 students	100 students
SPP Classrooms Cohort 3			160 students	160 students
SPP Classrooms Cohort 4				84 students
Total Students	**280**	**312**	**316**	**400**

By 2018–19, the Evaluation Team will be able to look at overall gains for children participating in SPP; compare gains between classrooms that have been in the program for one, two, and three years; and measure year-by-year gains for the first cohort of classrooms. The latter two comparisons are meant to capture the extent to which programs are getting stronger over time.

Overlayed comparative design using Randomized Controlled Trials (RCT)

A comparison group is critical to an assessment of SPP's impact on student learning and development because it allows DEEL to see which gains can be attributed to the SPP. Ideally, the comparison group is virtually identical to the group enrolled in SPP. Randomized controlled trials (RCT), which select the comparison group randomly, are regarded as the gold standard in research design. In an RCT, the analysis compares outcomes for children in the program to outcomes for children who are not in the program.[49] RCTs provide accurate estimates and require the smallest sample size, which make them less expensive to conduct. However, an RCT only works when there are many more children seeking entry to the preschool program than can be accommodated.

The best known preschool studies have used this approach.[50] The approach also creates a strong foundation for follow-up evaluation to assess lasting gains for children in Kindergarten and beyond. This approach provides the greatest confidence for answering well-defined questions about "what works."[51] It also provides the most precise estimates for any sample size, which is important because SPP begins as a relatively small program.

[49] Many of these children enter other preschool programs that may be of better than average quality, such as Head Start and private programs. These may not be typical of the experiences of all children eligible for Seattle preschool.
[50] Appendix A.
[51] Feuer, Towne, & Shavelson. (2002). Scientific Culture and Educational Research. Educational Researcher, Vol. 31, No. 8, pp. 4–14.

Children applying to the pre-K program are randomly selected from those children who apply in accordance with the SPP enrollment priorities.[52] Students who are not chosen by random assignment to participate in the program will be placed on a waiting list. This waiting list provides a comparison group. A more complicated version of this process can be applied if families apply for specific locations and rank their choices.

Table 8 shows this randomized trial design. The design parallels the design in Table 7: Pre-post design for impact analysis. The additional data collection will be done only for the comparison group, half of the number of children shown in Table 7. (Again, as more classrooms are added to the sample, fewer children are assessed in each classroom.) An equal number of children will be drawn from the waiting list.[53] This design will also allow comparisons of gains between classrooms in centers with different lengths of tenure in SPP. Note that Table 8 depicts the RCT for collecting evaluation data beginning in Year 2.

Table 8: Randomized Control Trial[54]

	2015–2016	2016–2017	2017–2018	2018–2019
	All students	*8 per classroom*	*4 per classroom*	*4 per classroom*
SPP Classrooms Cohort 1		112 students	56 students	56 students
Cohort 1 Control Group		112 students	56 students	56 students
SPP Classrooms Cohort 2			100 students	100 students
Cohort 2 Control Group			100 students	100 students
Total Students	**N/A**	**224**	**312**	**312**

Sample sizes could be smaller than shown for each cohort control group in Table 8 (if the waiting list is smaller). This design remains feasible even if the waiting list contains only half the number of children admitted to SPP. If SPP waiting lists are very small or nonexistent, the Evaluation Team has three alternatives for creating a comparison group. The best alternative is to obtain a demographically comparable sample of children in preschool and child care settings that are not yet part of SPP. The second best alternative is to use a regression discontinuity design (RDD). The RDD approach compares children who have just finished the SPP preschool to children who are just entering the SPP. This approach effectively controls for age by sampling children with birthdays just before and just after the cutoff date for entering preschool. For example, the Evaluation Team would sample children with July and August birthdays just finishing SPP to children with September and October birthdays just entering SPP. This design takes advantage of the basic randomness of birthdate relative to the school

[52] Note: in Year 1, Cohort 1 and Year 2, Cohort 2, students are assessed but not included in the RTC. See the Seattle Preschool Program Plan for more information: http://www.seattle.gov/Documents/Departments/OFE/AboutTheLevy/EarlyLearning/SPP_ProgramPlan2015-16_Final.pdf

[53] It is possible that not every center or classroom will have a waiting list; equal numbers are not required.

[54] Classroom Cohort 3 and Cohort 4 are not included in this design.

entry cutoff age, because it is a matter of chance whether a child's birthdate falls just before or after the cutoff.

RDD tends to find larger impacts than other designs because it estimates the full effects of the program on participants relative to those not attending. It also has strong claims to producing unbiased estimates. RDD requires a larger sample size, so the Evaluation Team recommends postponing this until fall of 2018.[55] Note that RDD does not provide a basis for longitudinal follow-up of impacts after children start Kindergarten. Therefore, the Evaluation Team would use RDD along with another approach that provides an SPP and comparison group for longer-term evaluation.

Using either an RCT design or RDD offers the best chance of accurately measuring program impacts on children's learning and development. An RCT design requires the smallest sample, which limits costs but presents the difficulty of locating children and families on the waiting list, persuading them to participate in the study, and making arrangements to collect their data at home or elsewhere. RDD requires only the participation of children admitted to the preschool program and, because all children are assessed at preschool or school, limits costs. RDD requires the assistance of the public schools in order for the Evaluation Team to collect post-program data on children as they enter Kindergarten.

If neither of these approaches are feasible, the remaining alternative is to construct a comparison group from a national dataset. This could be done using the Early Childhood Longitudinal Study-Birth Cohort (ECLS-B). However, this design is the least satisfactory of the alternatives because the comparison group may not be sufficiently comparable.

Classroom quality

The last component for the impact evaluation shifts the focus from the child onto the classroom. Linking classroom quality to children's learning and development increases confidence that the evaluation is accurately identifying the impact of SPP on those outcomes. To the extent that the evaluation finds links between increased classroom quality and increased gains in learning and development, this instills confidence that SPP is producing the desired results. The design for analyzing classroom quality follows the same pattern as the pre-post design.

Seattle should expect quality to vary among classrooms each year and to increase in classrooms as they participate longer in SPP.

Use

The Evaluation Team will conduct analyses for each of the research questions described in Section II (Align Evaluation Strategy with Program Standards). This section describes key examples that illustrate the types of analyses to be conducted.

- Compile simple descriptions of the children, families, and classrooms participating in SPP and in the comparison group to provide a basis for comparing participating children and families to the comparison group and Seattle's children and families in general.

[55] Although this approach has not been used in exactly this way, the Evaluation Team believes it could work well with a large enough sample size.

- Statistically model classroom quality overall, as well as classroom quality when controlled for program and staff characteristics (including the number of years classrooms have been part of the SPP, etc.) and the characteristics of the children and families they serve (including income, language spoken at home, etc.).
- Analyze progress for all SPP students each year and within each cohort of classrooms on various measures. Analyses will report how growth varies by child and family characteristics, as well as with classroom quality and comparing programs that came in earlier with programs that came in later.
- Compare gains for students in the SPP program to gains for children in a comparison group. The precise nature of these analyses depends on the SPP design. If a random selection process is used to form a waiting list comparison group, the statistical anlayses are straightforward (one group participated, one did not, this is how they differ). This comparison can be done for subgroups based on child and family characteristics (e.g., ethnicity, home language, parent education level, or child's prior experience with early care and education). Other designs depend much more heavily on statistical modeling.[56]

Who will conduct the analyses and timing

The Evaluation Team will conduct the impact analyses. Although the Evaluation Team will collect and analyze impact measurements starting in Year 1, there is no expectation that any progress toward the program outcomes will surface until the SPP providers have had time to implement program standards.

All of these analyses are conducted across years, but they become more complex as the sample grows each year with the addition of children and classrooms. Fortunately, they also become more robust with each year as the growing sample allows for better estimates of the impacts SPP is having on child outcomes. As classroom cohorts spend more time in SPP and make more progress on implementing SPP standards, the impacts for them can be expected to become stronger and easier to measure.

Process Evaluation

Because observed classroom quality, particularly the quality of teacher-student interactions, is correlated with improvements in children's learning and development,[57] results from analyses of process evaluations will be used to identify opportunities for improvements to classroom quality. For example, one exemplar program, New Jersey, began by analyzing data on program implementation, classroom quality, and child outcomes during the initial years of the Abbott preschool program. This information was used to focus statewide professional development and technical assistance on key areas that needed improvement.

[56] For example, use pre-test data for the comparison group and data from the family questionnaire to adjust for pre-existing differences between SPP and comparison groups.

[57] Reynolds, Arthur J., Temple, Judy A., Ou, Suh-Ruu, et al. (2007). Effects of a School-Based Early Childhood Intervention on Adult Health and Well-Being: A 19-Year Follow-Up of Low-Income Families, *Archives of Pediatric and Adolescent Medicine, 161(*8), 730–739.

The process evaluation, which will assess implementation of the SPP's design, provides a basis to determine how much progress providers and the SPP are making toward adhering to program standards. If the program model is not working as intended, process evaluation provides a way to identify and understand implementation challenges.

Process evaluation for the SPP will focus on activities and outputs that directly relate to the teacher-student interactions.

Use

The process analyses will be used to identify areas for improvement and measure progress toward adopting program standards. The process analyses will include data-driven theories about what aspects of the SPP are working well and what are not. Where possible, the Evaluation Team will work very closely with DEEL to understand important barriers to adopting standards, such as teacher retention, changes in site or program leadership, or other unanticipated factors. Although these natural occurrences that may hinder the implementation of the program are not explicitly listed in the tables that follow (Tables 9 and 10), they will factor into the analyses as necessary.

Table 9 lists each data source and describes the analysis and how it will be used.

Table 9: Data analyses and use

Program Area (Data Source)	Data Analysis	Use
Classroom Quality (CLASS and ECERS-3)[58]	• Analysis at both the classroom and provider level. • Comparison across providers.	• Identify patterns in the quality of teacher-student interactions and classroom environments. • Recognize areas of strength and weakness where best practices can be shared by DEEL with center directors or through other means.

[58] The Evaluation Team will anchor the analysis on CLASS and ECERS data collected by the Evaluation Team but will opportunistically consider EA and coach-collected data where it can augment the analysis.

Program Area (Data Source)	Data Analysis	Use
Curriculum Implementation (DEEL Self-Evaluation Tools)[59]	• Descriptive information about curriculum and training. • Analysis of implementation by teacher and coach assigned. • Curriculum training and implementation fidelity analyzed for trends and insights into areas where children may not be getting as deep or broad an exposure to the instruction required for quality or improved outcomes.	• Determine if curriculum is implemented fully and correctly. • Assess relationship between curriculum implementation and coach support.
Early Learning Standards (TS GOLD)[60]	Report the implementation of early learning standards.	Determine if the early learning standards are all being implemented effectively.
Teacher Credentials and Professional Development Professional Development (WA DEL MERIT Database; Teacher Survey)	• Track teacher credentials and educational attainment. • Teacher professional development attainment.	• Assess whether the teaching workforce has the proper credentials to succeed. • Determine if teachers are obtaining training to improve teaching.
System Evaluation (Director Surveys)	• Analysis of time spent supporting teachers, the classroom environment, and quality learning. • Analysis of whether program policies/requirements are helping or impeding the implementation and improvement.	• Determine how to leverage leadership to effectively implement program at the site level. • Assess improvement strategies and necessary next steps to more effectively implement the program standards.

[59] Fidelity of implementation of curriculum is significantly more difficult to assess and would require trained experts as time and budget allow. The Evaluation Team will determine the appropriate level of process evaluation in this area and will consider deploying experts to assess curriculum fidelity as the program evolves.

[60] Data collection is predicated on the assumption that there are clear program standards aligned with TS GOLD items.

Program Area (Data Source)	Data Analysis	Use
System Evaluation (Teacher Surveys/Interviews)	• Qualitative assessment of how a teacher is effectively incorporating the use of data and integrated systems. • Analysis of whether program supports are helping or impeding the implementation of the program.	• Determine how to leverage teachers to effectively implement program at the classroom level. • Assess improvement strategies and necessary next steps to more effectively implement the program standards.
System Evaluation (Coach Surveys/Interviews/ Focus Groups)	• Qualitative assessment of how a coach is effectively helping both the director and teachers to effectively incorporate the use of data and integrated systems. • Analysis of whether program supports are helping or impeding the implementation of the program.	Determine how to leverage coaches to effectively implement program standards at the classroom level.

Who will conduct the analysis and timing

Given that the SPP will grow from approximately 14 classrooms in 2015–16 to a target of 100 classrooms in 2018–19, the evaluation will provide a snapshot of each cohort of providers as they progress.

The Evaluation Team will provide an independent, third-party assessment of key teacher-student interaction data elements described in Table 10.

In Year 1, the Evaluation Team will analyze the 2015–16 cohort of providers with a focus on quality improvement, curriculum implementation, progress toward teacher credentials, and other process evaluation indicators to establish an independent baseline. The Evaluation Team will compare all future analyses of the 2015–16 cohort to its baseline.

In Year 2, SPP will add a larger cohort, with new classrooms, centers, and teachers supported by a growing set of coaches and staff at DEEL. The Evaluation Team will conduct an analysis to establish an independent baseline for the 2016–17 cohort. The Evaluation Team will continue to establish independent baselines for each subsequent cohort.

Years 1 and 2 will focus on descriptive analysis of basic process evaluation indicators related to teacher-student interactions. In Year 1, the Evaluation Team will produce process evaluation results to assess at minimum classroom quality (CLASS and ECERS-3), curriculum implementation, teacher qualifications and credentialing, and early learning standards. In Year 1, the Evaluation Team will observe providers in Cohort 1 and develop a process evaluation model that is scalable in later years. The Evaluation Team will use Year 1 and potentially Year 2 as a pilot to develop analytics and align data with analytic goals.

Years 3 and 4 of the program will have 14 to 15 classrooms that have had three years in the program and two to three larger classroom cohorts that are newer to SPP. The process evaluation will become more robust as more data are available and SPP's systems and processes become more established. The Evaluation Team will provide the same descriptive statistics as in earlier years as well as longitudinal and more robust statistical analyses over time.

By Year 3, as SPP matures, the 14 classrooms that formed the first SPP cohort will be in a position to realize improvements. With clean and accurate data available, the Evaluation Team will provide statistical analysis for each element of the classroom measurement instruments, TS GOLD scores (to measure adherence to program standards), professional development hours, and educational attainment.

Table 10 shows the expected targets for each cohort of classrooms by the end of each program year.

Table 10: Expected process targets for SPP program area by year of classroom participation

Program Area (Data Source)	Expected Targets in Year 1 of Classroom's Participation in SPP[61]	Expected Targets in Year 2 of Classroom's Participation in SPP	Expected Targets in Year 3+ of Classroom's Participation in SPP
Classroom Quality (CLASS and ECERS-3)[62]	• Confirm that all SPP classrooms have obtained CLASS scores at minimum thresholds (3.5 for CLASS ES and 2.0 for CLASS IS). • Confirm that all SPP classrooms have obtained ECERS-R scores at minimum thresholds of 3.0.	• CLASS scores are improving beyond minimum thresholds. • ECERS-R scores are improving beyond minimum thresholds. • Analysis will include progress on CLASS dimension scores relative to targets and relative to the baseline.	• CLASS scores are materially improving toward standards given current trendlines (6.0 for CLASS ES and 4.5 for CLASS IS). • ECERS-R scores are materially improving toward standards given current trendlines (5.0).
Curriculum Implementation (DEEL self-evaluation Tools)[63]	• SPP approved curriculum has been implemented in each classroom, and teachers have been trained on the curriculum.	• Teachers receive curriculum training and are reporting use of curriculum.	• All teachers are implementing the curriculum with fidelity.64

[61] Please note: Year 1 of program implementation is the first of SPP program implementation in the classrooms that are being measured.
[62] The team will anchor the analysis on CLASS and ECERS data collected by the Evaluation Team but will opportunistically consider EA and coach-collected data where it can augment the analysis.
[63] The Evaluation Team will determine the appropriate level of process evaluation in this area of assessing fidelity of implementation and will consider deploying experts to assess curriculum fidelity as the program evolves.
[64] The method to evaluate will need to be determined once the program has been launched, initial results are analyzed, and the level of resources available to deploy curriculum experts is appropriate.

Program Area (Data Source)	Expected Targets in Year 1 of Classroom's Participation in SPP[61]	Expected Targets in Year 2 of Classroom's Participation in SPP	Expected Targets in Year 3+ of Classroom's Participation in SPP
Early learning standards (TS GOLD)[65]	• Early learning standards are understood and implemented by teachers as reported by DEEL and coaches.	• Student TS GOLD scores are increasing in the areas that are aligned with early learning standards (data provided by DEEL). Similarly, analysis of alignment between children's scores on TS GOLD and early learning standards will reveal gaps that could impact outcomes.	• Early Learning Standards: Student TS GOLD scores are increasing in the areas where there are program standards and gaps identified in Year 2 have been mitigated.
Teacher Credentials and Professional Development (WA DEL MERIT Database; Teacher Survey)	• Seattle has an accurate picture of the educational attainment of all teachers. This assumes all teachers have completed their professional development achievement awards in MERIT and that WA DEL provides the Evaluation Team with clean and timely data.	• Increasing numbers of teachers are taking advantage of tuition and advisory services to obtain BA degrees, and the SPP workforce is adequately progressing toward program standards. • Analysis of professional development and teacher credential data in MERIT and teacher interviews or surveys will include identification of gaps in teacher preparation and/or progress toward targets.	• The number of teachers with or working toward BA degrees and ECE credentials is increasing and, given trendlines, teachers are forecasted to materially improve toward the standards set by the professional development requirements.

[65] Predicated on assumption there are clear program standards aligned with TS GOLD items.

Program Area (Data Source)	Expected Targets in Year 1 of Classroom's Participation in SPP[61]	Expected Targets in Year 2 of Classroom's Participation in SPP	Expected Targets in Year 3+ of Classroom's Participation in SPP
System Evaluation (Director Surveys, Teacher Surveys/Interviews, Coach Surveys/Interviews/ Focus Groups)	• Directors and teachers are using information and coaching support to improve their practices. • Coaches are using information from the evaluation as well as the self-evaluation rubric to guide their coaching.	• Directors are using information to support and improve the supporting environment. • Coaches are using information from their own classroom observations to inform their work with teachers. • Teachers are using information to improve their practices. • Systems analysis will include levels of use of data and/or communication with comparison across providers to accelerate progress for the program as a whole.	• Directors are improving their use of information to support and improve the supporting environment. • Coaches are improving their use of information from their own classroom observations to inform their work with teachers. • Teachers are improving their use of information to improve their practices. Improvement will be measured by longitudinal results as captured in the survey design.

Self-evaluation

Analysis of the site-level assessment rubric will allow DEEL to determine if SPP is materially improving toward the benchmarks of high-quality programs as described in the program standards. In partnership with its providers, the New Jersey program team analyzed results from the self-evaluation tool and used the results to generate ideas and approaches for improving classroom environments and program elements to support child outcomes.[66] Thoughtful analysis of New Jersey's self-evaluation results generated meaningful improvements to program quality over the first decade of the Abbott preschool program.

The first analysis, based on the data from the continuous self-evaluation, will create a benchmark or a standard to set necessary improvement. DEEL can conduct further analysis at different intervals depending on what is being analyzed. For example, program hours are unlikely to change during the school year and therefore can be assessed once a year. Conversely, the way teachers implement new curriculum may improve over the course of the school year and should be analyzed periodically at DEEL's discretion.

[66] Interview with Ellen Frede, PhD, 6/17/2105.

Use, who will conduct the analysis and timing

DEEL staff will use the analysis to determine the quality of SPP implementation. As they were in New Jersey, these findings will be used to drive improvements and learn how to best implement a program that leads to positive outcomes for students.

DEEL will conduct the analysis of the data from the self-evaluation instrument(s). The analysis will be customized to the needs of DEEL staff. DEEL will use an analysis of the site-level assessment rubric to work with sites until SPP targets are met.

The Evaluation Team will incorporate elements of the self-evaluation, as supplied by the city of Seattle, into a dashboard to create an ongoing picture of provider needs and coaching activities.

Data by provider will be aggregated, analyzed, and shared internally on a recurring and regular basis. This output will allow program leadership and staff to observe performance levels at a point in time but also to see annual trends and track momentum.

Section V: Implement Improvements

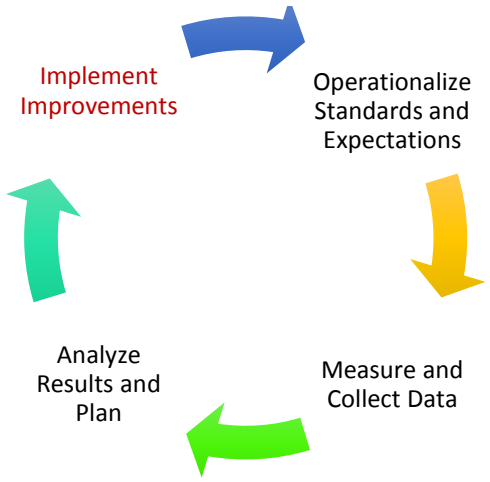
This section describes the ways that DEEL will implement improvements at the classroom, provider, and programmatic level.

SPP Evaluation's Approach to Implementing Improvements

SPP includes three strategies for identifying and implementing improvements:

- Coaching for providers and teachers, to address improvements at the provider and classroom level
- A Peer Learning and Improvement Network comprising providers and DEEL to address improvements at both the provider and programmatic level
- A Leadership Team at DEEL to address improvements at the programmatic level

Figure 5: CQI Cycle for SPP - Part IV



Program improvements through coaching

The coaching model used by DEEL, which is designed to support teachers in their development and improve program quality, is rooted in best practices from other preschool programs.[67] The coaching model is a relationship-based approach to implementing improvements. Coaches, along with center

[67] Skiffington, S., Washburn, S., and Elliott, K. (May 2011). Instructional Coaching: Helping preschool teachers reach their full potential. *Young Children*.

directors, who understand the unique circumstances of each teacher and classroom, are able to contextualize the evaluation analyses and identify the improvements needed to make progress toward the desired results.[68] For example, the CLASS assessment of an individual teacher may suggest that certain aspects of teacher-child interactions need improvement. The coach and director, armed with both the assessment analyses and their own impressions from observing the teacher in action, are in the best position to identify specific strategies the teacher can use to improve his or her interactions with students.

Coaches have expertise in early childhood education and will work directly with providers—specifically center directors and teachers. Coaches will work with directors to assess, analyze, identify, and implement improvements frequently throughout the school year.

Coaches, in partnership with providers, will develop Quality Improvement Plans (QIPs) for providers and classrooms based on an analysis of self-evaluation data. Each QIP will outline specific recommendations to advance the learning environment and classroom practice. DEEL coaches will use the QIPs to help directors and teachers implement improvements and track their progress.

Coaches are in the best position to translate program theory, measurement, and results into practical strategies and applications that directors and teachers can use to improve professional practice. The coach is also a critical bridge between individual providers and DEEL's policymakers; coaches will have on-the-ground exposure to which SPP policies work and which do not work in practice and can advise DEEL accordingly.

Peer Learning and Improvement Network

Peer learning is a powerful and effective approach to implementing improvements because good ideas become better ones when a diverse group of practitioners discusses and refines them based on practitioners' own expertise and experiences.[69] Furthermore, because directors and teachers have a key role in identifying improvements and developing strategies for implementing them, they are more invested in the success of these ideas.[70]

It is important to set the right tone to maximize these benefits. Participants must explicitly understand that their charge is to support one another in working toward quality. Meetings need to be a safe space where providers can confront their challenges honestly and openly and work together to wrestle with solutions.[71] If the funding agency is involved, it should engage as a collaborative partner, not a compliance monitor. For this reason programs often hire a third-party facilitator to run the peer group meetings.[72]

---

[68] Isner, T., Tout, K., Zaslow, M., Soli, M., Quinn, K., Rothenberg, L. & Burkhauser, M. (2011). Coaching in early care and education programs and Quality Rating and Improvement Systems (QRIS): Identifying promising features. Report produced for the Children's Services Council of Palm Beach County. Washington, DC: Child Trends.
[69] Interview with Ellen Frede, PhD, 6/17/2015.
[70] Ibid.
[71] Ibid.
[72] Ibid.

DEEL will convene a working group comprising participating center directors and other key stakeholders called the Peer Learning and Improvement Network (Peer Network). The Peer Network will be a forum for providers to share information, learn from each other, discuss the areas they need to improve, and brainstorm strategies for implementing improvements. The Peer Network will also advise DEEL on strategies for making improvements at the programmatic level.

The Peer Network will oversee development of the continuous self-evaluation tool used by providers and described in Section III (Measure and Collect Data).

Early Learning Leadership Team at DEEL

DEEL will analyze and use findings from the self-evaluation and third-party evaluation analyses to identify and implement program improvements. The group responsible for this will be an internal SPP Leadership Team comprising the director of early learning and leads for policy and planning, quality improvement, operations, communications, and data. Meetings will be regularly scheduled.

High-quality public preschool programs use leadership teams to ensure that the program standards are integrated with each other for successful outcomes. A culture of high-expectations (with support) exists throughout the system, and stakeholders are included when appropriate.[73]

The Early Learning Leadership Team has two key responsibilities:

- Identify and implement programmatic improvements based on the self-evaluation, the process evaluation, and the outcome evaluation
- Plan and implement stakeholder engagement and communication based on the self-evaluation, the process evaluation, and the outcome evaluation

Roles and Responsibilities

In all four steps in CQI, DEEL, coaches, providers, and the Evaluation Team have distinct roles and responsibilities. Nowhere in this CQI process are clear roles and responsibilities more important than in implementing improvements. This section describes roles and responsibilities for this phase of the CQI.

Identifying and implementing improvements to SPP and individual providers

The DEEL Early Learning Leadership Team will be responsible for using analyses from the self-evaluation, process evaluation, and, after the program is successfully implemented, impact evaluation to identify and implement improvements. For example, if the process evaluation suggests that many providers are not meeting standards related to teacher-child interactions, the DEEL Leadership Team may decide to adjust policies and/or redirect resources to support professional development around teacher-child interactions.

[73] Minervino, Jim (with contribution from Robert C. Pianta, PhD, University of Virginia). Lessons from Research and the Classroom: Implementing High-Quality Pre-K that Makes a Difference for Young Children (White Paper, Bill & Melinda Gates Foundation, September 2014).

The Evaluation Team will support DEEL Leadership Team in identifying and implementing improvements by producing and submitting analysis and reports at regular intervals, as follows:

- Process evaluation status updates each November and interim reports each February and May. The status updates will consist of a brief memo that outlines progress and highlights, including, for example, teacher training and educational attainment. The interim reports will provide the DEEL Leadership Team with real-time information that DEEL can use to make course corrections to SPP.
- Impact evaluation status updates in November, February, and May. These status updates for the DEEL Leadership Team will include a brief memo and discussion of progress on the impact evaluation. This memo will provide a brief overview of progress on classroom observations and student assessments.
- After receiving a status update or interim report, DEEL may want to request additional custom analytics to gain more insights. The budget includes hours for developing custom analytics.
- A comprehensive process evaluation report and a comprehensive impact evaluation report each August. Each report will summarize the results of the analyses described in Section IV (Analyze Results and Plan). These reports will be technical in nature and will reflect the format, style, and voice that is appropriate for each evaluation type.
- An executive summary of the annual reports in a consistent format, style, and voice. The executive summary will include a synthesis of annual analyses. DEEL will be able to adapt this executive summary for communications with other stakeholders.
- A secure transfer of all raw data used for evaluation purposes.[74]

Communicating analyses

How the analyses are communicated to stakeholders is just as important to an evaluation as clear standards; reliable, valid, and accessible data; and a well-designed analysis. According to the research, successful evaluations communicated findings on a regular basis to help providers, program leaders, and other stakeholders make use of results to move toward defined quality benchmarks and evidence-based practices.[75]

The DEEL Leadership Team and DEEL's Communications Team will be responsible for communicating evaluation analyses and improvements (including the timing, messaging, and format) to the following stakeholders:

- Coaches and education specialists, to convey how they are working with teachers and providers
- Individual providers, to use the information to identify and implement improvements
- The Seattle Mayor's office, city council members, the Levy Oversight Committee, and other city leadership to understand and use it to inform their decision making

[74] The term "raw" in this case means that the Evaluation Team will not alter the data in any way beyond what it has already done for purposes of analysis and reporting.

[75] Metz, A., Naoom, S. F., Halle T., & Bartley, L. (May 2015). An Integrated Stage-Based Framework for Implementation of Early Childhood Programs and Systems (OPRE Research Brief #2015–48, p. 11).

- Key partners such as the Department of Early Learning (DEL) and Seattle Public Schools (SPS)

Communicating with providers

In most evaluations, the early self-evaluation and process evaluation data were used to make mid-course corrections and increase the quality of implementation. These early data were intended for internal audiences: those responsible for program improvement. If stakeholders do not understand these, they cannot use them to make improvements. Therefore, early analyses need to be contextualized by someone like a coach, who both understands the evaluation results and the providers who will need to use them to make improvements.

Communicating with other stakeholders

DEEL, which will be responsible for identifying and making program improvements based on the evaluation analyses, is best positioned to decide when and how to communicate these improvements to the various stakeholders. When DEEL communicates with stakeholders such as the mayor, Seattle City Council, the Preschool Levy Oversight Committee, and the public, it will want to share not just the analyses of the evaluation but how those analyses informed decisions about improvements. That way, DEEL can both celebrate SPP's successes and progress and proactively demonstrate what it is doing to continuously improve and address program elements that are not yet yielding desired results.

A key lesson learned from the research is that public reports on evaluations should not be circulated to external audiences until the preschool program has had time to make progress. Care must be taken that early analyses are not circulated in ways that would publicly embarrass or shame participating providers. This is a particular concern when there are a small number of participants and it would be easy to single out providers. If providers believe there is a risk that early evaluation data on their programs may be circulated widely, they may be reluctant to participate in the evaluation and provide accurate and complete information. Consequently, the quality of the data collected will be undermined, making it much less useful for program improvement. After several years, when providers have had an opportunity to improve, preschool programs can make data that demonstrate progress against the baseline available more widely.

If there is concern that evaluation analyses may be misused or misinterpreted in service of an agenda other than achieving positive child outcomes, the DEEL Leadership Team may need to carefully plan how and when it disseminates evaluation analyses to the external audiences.

Role of the Evaluation Team in communications

When performing evaluation, the credibility of the evaluation rests on the Evaluation Team's ability to retain its status as a neutral third party. As such, there can be no perception that anyone with specific interests in SPP's success is involved in developing the annual or interim evaluation reports and analyses. Similarly, while it may be completely appropriate and desirable for the city of Seattle to tailor messaging about evaluation and improvements to stakeholders, it would be inappropriate for the Evaluation Team to do so. Any tailoring or reshaping of messages could be misinterpreted as bias and compromise the Evaluation Team's neutrality and credibility, which would compromise the evaluation's neutrality and credibility.

The Evaluation Team will support the DEEL Leadership Team's efforts to communicate results by making Evaluation Team members available to present and discuss its findings to key stakeholders when necessary. In these circumstances, the Evaluation Team will communicate analyses in an objective and independent manner. This means that it will not provide communication materials to the city of Seattle in advance to edit language, iterate drafts, tailor messages, or otherwise engage the city of Seattle in a manner that would compromise its status as a neutral third party.[76]

Section VI: FCC Pilot Evaluation Development

The SPP Action Plan states, "After initial program start-up the City will develop a Family Child Care (FCC) Pilot program to assess whether, and how, partnerships with FCC providers can be implemented to achieve the same quality standards attained by center- and school-based providers, in a cost-effective manner." This section describes the process through which the FCC pilot will be developed and how the evaluation will be designed.

About Family Child Care

Family Child Care (FCC) is an integral part of child care services. In Seattle, approximately 23 percent of all children in licensed child care are in FCC, while the remaining 77 percent are in child care centers.[77,78] Because FCCs may offer child care in addition to preschool services, families can often use one provider for all of their child care needs. FCC care also offers families the opportunity to select a provider who shares their cultural norms and values. FCC providers and the children and families they serve are generally more representative of communities of color and those whose primary language is other than English.

Given the importance of FCC providers to a comprehensive child care system, it is important to study how they can be integrated into the SPP. As such, to fully determine whether FCC providers who meet all SPP criteria can produce comparable results to center-based sites, SPP will include FCC pilot evaluation. The Evaluation Team will design the FCC Pilot Evaluation in 2015–16 for implementation in 2016–17.

As with the center- and school-based providers, DEEL will be responsible for conducting outreach and recruiting FCC providers to participate in the SPP, determine provider and participant eligibility, manage monitoring and compliance, and provide a variety of supports to FCC participants.

[76] The Evaluation Team will provide its analysis and results to the city of Seattle in advance so that the city has the opportunity to digest the team's findings, ask questions, etc.

[77] BERK Consulting (2014) *Recommendations for Seattle's Preschool for All Action Plan.* Available at: http://www.seattle.gov/Documents/Departments/OFE/AboutTheLevy/EarlyLearning/BERK_Recommendations.pdf

[78] Licensed child care represents only a portion of all child care. Providers who care for children for four hours or less are not required to be licensed.

FCC Pilot Evaluation Overview

The Evaluation Team will begin the design phase of the FCC Pilot Evaluation in the fall of 2015; the final evaluation plan is due in the spring of 2016.

The FCC Pilot Evaluation will:

- Assess whether the SPP comprehensive evaluation strategy is applicable in FCC settings
- Make recommendations to DEEL about adjusting the strategy as needed to address the opportunities and constraints presented by FCC contexts

To develop the FCC Pilot Evaluation Plan:

1. The Evaluation Team will conduct a literature review on the few preschool programs that incorporate FCC providers.
2. The Evaluation Team will conduct interviews with local leaders representing diverse perspectives on FCC.
3. DEEL, with support from the Evaluation Team, will create an Advisory Committee comprising representatives from key stakeholder groups.

Outreach to ensure participation in the FCC Pilot Evaluation will be critical. By involving key FCC stakeholders in the design process from the beginning, the Evaluation Team will build buy-in among members of the FCC provider community.

Design Step One: Review of FCC models of high-quality early education

While FCC providers are common, there are only a few examples of large-scale preschool programs using FCC providers. In 1995, Head Start recognized FCC as a viable option for delivering its services. In 2008, Head Start established Head Start Program Performance Standards (HSPPS) for FCCs. In 2014, the WA DEL made application for state-funded preschool slots open to licensed FCC providers participating in EA.

The Evaluation Team will gather information on FCC models delivering high-quality programming—especially those that are integrated into city-funded pre-K programs—to ensure the FCC Pilot Evaluation captures lessons learned. This research will include what works and what does not work and create a basis for the construction of the pilot.

Design Step Two: Interviews with local leaders representing diverse perspectives on FCC

The Evaluation Team will also work with DEEL to identify and interview local leaders representing diverse perspectives on FCC, such as Puget Sound Educational Service District (PSESD), which is one of nine regional educational agencies created by the Washington legislature that provides support services for early learning and K-12 providers; Child Care Resources, a nonprofit organization supporting early learning providers; Service Employees International Union (SEIU), the union that represents FCC providers; and the key leaders from the FCC provider community.

The purpose of these interviews is to understand the local FCC landscape and how FCC providers deliver or can deliver high-quality preschool services, to surface issues critical to the FCC community that need to be addressed in the FCC Pilot Evaluation design, and to consider the needs and interests of the FCC community.

Design Step Three: FCC Pilot Evaluation Advisory Committee

The Evaluation Team and DEEL will identify Advisory Committee members who are able to represent their community's interests, provide meaningful feedback on the design of the FCC Pilot Evaluation, and help create buy-in among FCC providers who are likely candidates to participate in the FCC Pilot Evaluation. The Advisory Committee will include representatives from key stakeholder groups, including but not limited to: DEEL, SEIU, and key partners from the FCC community. The Advisory Committee will be representative of the diversity of race, language, and culture in Seattle's FCC community and include leaders from immigrant and refugee communities. The Advisory Committee will be composed of no more than 15 members in order to be productive over the short time frame for its work.

The Evaluation Team will work with DEEL to determine the precise structure of the Advisory Committee. For example, to ensure that everyone on the Advisory Committee has an equal contribution to the deliberations, including participants who come from language or cultural groups that do not have a tradition of speaking up in larger meetings, it may make sense to create a structure with defined sub-groups and/or simultaneous small group work, instead of a single Advisory Committee. The Evaluation Team will also work with DEEL to determine the best balance between time needed to delve deeply enough into the issues and asking for reasonable time commitments from members, many of whom will be providers. The Evaluation Team will carefully construct agendas to maximize and focus member input.

To meet the time lines for delivering a final plan for the FCC Pilot Evaluation, the Evaluation Team will need to convene the Advisory Committee three times, beginning in November 2015. The goal for the Advisory Committee is to provide options and recommendations to DEEL for the structure of the FCC Pilot Evaluation, which will be submitted in the spring of 2016.

FCC Pilot Evaluation Design: Guiding Principles

The FCC Pilot Evaluation design will be guided by two key principles:

1. To the maximum extent possible, the FCC Pilot Evaluation will match the structure, expectations, and supports of SPP as it is rolled out in centers and schools. This means FCC providers will need to meet similar standards to center-based providers; and they will receive coaching and the same level of feedback from evaluation results as center- and school-based providers.
2. The FCC Pilot Evaluation Plan will adapt the SPP Evaluation Strategy to reflect the unique needs and realities of delivering early learning in a home-based setting.

Evaluation of the FCC Pilot will be incorporated into the overall comprehensive evaluation of SPP as described in this Strategy. Thus, the Evaluation Team will need to ensure, where possible, that measurement instruments used in the SPP Evaluation Strategy are compatible with FCC settings or can

be easily and effectively adapted to the FCC setting where feasible. The Evaluation Team will need to adapt FCC versions of the environment assessments. The Advisory Committee will advise on these adaptations.

Section VII: Four-Year Budget for External Evaluation

This section presents the projected four-year budget for portions of the evaluation that:

- Begin in the 2015–16 school year and
- Are managed by the External Evaluation Team

The aspects of the SPP Evaluation Strategy that will be overseen by DEEL, such as self-evaluation and analyzing parts of the site-evaluation rubric, are not described in this section.

The adopted four-year budget for impact and process evaluation as described in this report is $2,045,414, as depicted in Table 11 below:

Table 11: Adopted budget for impact and process evaluation

Adopted Budget: External Evaluation	SY 2015–16	SY 2016–17	SY 2017–18	SY 2018–19	Total
Total Expenses	$669,906	$436,666	$459,979	$478,863	$2,045,414

This budget averages $511,354 annually, with more expenses planned in Year 1 (2015–16), followed by lower expenses in Year 2 (2016–17), and a slight increase in spending in Years 3 (2017–18) and 4 (2018–19). DEEL's adopted budget is reasonably close to national benchmarks and therefore should be sufficient to execute the Comprehensive Evaluation Strategy.[79] Please note that the amount for the Family Child Care Pilot Evaluation is addressed later in this section and not included in Table 11.

Budget Overview

The budget for impact and process evaluation includes three major expense categories:

1. Development of the Comprehensive Evaluation Strategy
2. Process Evaluation
3. Impact Evaluation

The cost to develop the SPP Evaluation Strategy, including the Technical Report for the City of Seattle (a literature review) was budgeted at $148,726. The budget for Process Evaluation is $869,228, and the budget

[79] Given that evaluation is based on a complex set of variables that evolve in parallel with the program, the Evaluation Team also recommends that DEEL engage in philanthropic partnerships to augment its research and evaluation activities where warranted and feasible. By doing so, DEEL will be in the strongest position possible to assess and evaluate its program for the benefit of the children of Seattle. These efforts and associated funds are not included in this budget section.

for Impact Evaluation is $1,027,460. Refer to Table 12 for a line-item, four-year budget for method of evaluation. A description of each budget category follows Table 12.

Table 12: Four-year budget for Process Evaluation and Impact Evaluation

Adopted Budget: External Evaluation	SY 2015–16	SY 2016–17	SY 2017–18	SY 2018–19	Total
Evaluation Strategy	$148,726	$0	$0	$0	$148,726
Process Evaluation					
System Evaluation	$51,000	$51,000	$51,000	$0	$153,000
Curriculum Implementation	$20,000	$23,000	$14,000	$0	$57,000
Professional Development	$29,000	$31,000	$37,000	$0	$97,000
Classroom Environments	$49,000	$66,000	$64,000	$33,040	$212,040
Program Standards	$27,000	$31,000	$37,000	$38,000	$133,000
Program Management	$38,403	$38,300	$53,500	$30,926	$161,129
Subtotal	$214,403	$240,300	$256,500	$101,966	$813,169
Contingency (7%)[80]	$15,753	$19,253	$21,053	$0	$56,059
Total Process evaluation	**$230,156**	**$259,553**	**$277,553**	**$101,966**	**$869,228**
Impact Evaluation					
Analysis & Management	$63,729	$130,657	$134,017	$138,404	$466,807
Observations & Assessments	$47,459	$115,613	$138,650	$162,151	$463,873
Travel	$2,860	$4,554	$5,585	$6,640	$19,639
Supplies/Other	$2,642	$23,503	$24,933	$26,062	$77,141
Total Impact Evaluation	**$116,691**	**$274,327**	**$303,185**	**$333,258**	**$1,027,460**
Subtotal External Evaluation	$495,573	$533,880	$580,738	$435,223	$2,045,414
Carry-Forward	$174,334	($97,214)	($120,759)	$43,639	$0
External Evaluation Budget	**$669,906**	**$436,666**	**$459,979**	**$478,863**	**$2,045,414**

[80] The contingency fee represents 7% of the total 4-year budget and is distributed annually based on scope of work

Budget Description by Major Expense Categories

The Evaluation Team developed the budget based on an estimate of time and materials, as well as its own experience conducting similar evaluations. Both process and impact budgets are "bottoms-up" budgets; in other words, they include estimates of specific activities over the course of the four years to arrive at a total. A line-item explanation by evaluation activity is as follows:

Support the development of the SPP Evaluation Strategy

The budget to support the development the SPP Evaluation Strategy will have already been expended by the time of strategy publication.

Process evaluation budget

For each process evaluation budget component, the Evaluation Team estimated costs for specific activities and for three separate stages of the evaluation: 1) development of the tool, survey, or analysis; 2) execution of the tool or data collection; and 3) analysis and reporting of the results. Generally, costs for development, analysis, and reporting are a function of the Evaluation Team's time and effort, while costs for execution is linked to the number of teachers, directors, coaches, classrooms, or children in the program. Execution of data collection often involves interviews, surveys, or primary data collection where cost is driven by the program scale. Even in the case where execution involves requesting, cleaning, and managing data from existing sources, the scale and scope of the data impact the costs of the data management activities.

- *System Evaluation* consists of the cost to design, administer, and report interviews and surveys of program directors, teachers, and coaches. Costs would normally rise for this activity year over year due to increasing program volumes in providers and classrooms, but the Evaluation Team plans to offset this effect either by using online surveys or reverting to surveying a sample of individuals in later years or both. For the first year, the Evaluation Team will perform more in-depth discussions with provider staff to set up survey design and take advantage of the relatively small number of program participants. For the last year it is anticipated that DEEL will be positioned to take over this work and perform a similar activity internally as necessary.
- *Curriculum Implementation* represents evaluation to measure whether and how well the approved curriculum is being implemented in SPP classrooms. This budget includes a checklist intended for each classroom indicating whether the curriculum has been purchased and whether teacher and director training on the curriculum has occurred. The Evaluation Team will monitor which providers select each of the two pre-approved curricula, HighScope or Creative Curriculum, and attempt to correlate these choices to overall provider and classroom performance as the budget allows. Please note that the Evaluation Team will not be able to conduct a comprehensive curriculum fidelity evaluation because this type of evaluation is prohibitively expensive and involves national curriculum experts to assess curriculum execution and usage on-site. For the last year it is anticipated that DEEL will be positioned to take over this work and perform a similar activity internally as necessary.

- *Professional Development* involves importing, cleaning, analyzing, and reporting data from MERIT operated by WA DEL. The purpose of this activity is to analyze whether teachers either meet or are sufficiently progressing toward degree and credential program requirements. Please note that this budget assumes that all teachers register themselves in MERIT and update professional development information annually and that WA DEL continues to adequately support MERIT by investing in system maintenance, system upgrades, and data quality. For the last year the Evaluation Team will provide the raw data from MERIT to DEEL, and an internal data team will perform process evaluation internally as necessary.
- *Classroom Environments* includes analysis of the data at the dimension level to determine progress toward program standards. This activity will also validate that all SPP classrooms maintain CLASS and ECERS scores at minimum thresholds, achieve or are progressing toward CLASS and ECERS program goals at the appropriate time, identify performance gaps, and forecast future trendlines based on current results. Please note that this category does not include data collection; the Evaluation Team will rely on data collected from the field as described in Table 12 above (in the "observations and assessments" line item in the impact evaluation component of the budget). For the last year it is anticipated that DEEL will be positioned to begin taking over this work and perform a similar activity internally; the Evaluation Team has budgeted funds in this category to provide technical assistance and ensure a smooth transfer of work and analyses.
- *Early Learning Program Standards.* The Evaluation Team will assess the alignment of Early Learning Standards with student learning through TS GOLD scores. The budget assumes that DEEL will collect and share a clean and timely TS GOLD dataset annually with the Evaluation Team.
- *Program management* provides funds for the prime consultant to manage, coordinate, and administer all evaluation activities between the prime consulting team and all subconsultants. In addition it also includes limited funds for:
    a. Participating in presentations of evaluation results to city of Seattle executives and other key stakeholders such as community-based organizations, SPS, and WA DEL.
    b. Customized data-driven targeted analysis. Examples could include the development of specific, targeted analytics to elucidate specific subsegments of the SPP.
    c. Evaluation Team time to provide a secure raw data transfer to DEEL.

  To provide DEEL with the flexibility to request customized analytics, this line item will be allocated on a time and materials basis until funds in this task are exhausted.
- *Contingency:* Evaluation budgets vary based on a variety of factors, including (but not limited to) study design, number of children, type of instruments used, availability of data, length of time, and numerous other factors. A wide range of variables outside the evaluator's control influence the cost of process evaluations, including the maturity and stability of a program; where and how it is being implemented; the capacity, experience, and expertise of program implementers; and demographic and socio-economic characteristics of program participants. To provide flexibility to respond to these conditions, there is a contingency line item of 7 percent% of the process evaluation subtotal.

Impact evaluation budget

- *Analysis and Management* represents resources to design, analyze, report, and manage the Impact Evaluation. Expenses include funds to support principal investigation and impact design; establish protocols for informed consent and assessments, field protocols and manuals, IRB protocols; develop surveys for the collection of demographic and socio-economic information; analyze qualitative and quantitative data; and support training, field supervision, and project management and coordination; and develop reports and status updates for DEEL. This budget category primarily consists of time and materials of NIEER staff to manage, execute, and report Impact Evaluation. Salaries, fringe, and indirect costs are included in this estimate.
- *Observations and Assessments* includes funds to support classroom observations (CLASS and ECERS) and child assessments. Activities include providing training for assessors; coordinating and carrying out the data collection, reliability checks and quality control; and data cleaning and entry. A budget for child assessments is also included for four children per classroom on a subset of the Woodcock-Johnson, Peabody Picture Vocabulary Test IV, and a set of Executive Function assessments. These assessments will be conducted pre- and post-testing in English and in Spanish (accommodations for children whose primary language is not Spanish or English will also be made even if the assessments are not available and normed in other languages). The budget also includes a limited budget to oversee and administer observations and assessments. This budget line item is driven by the volume of classrooms and children assessed and observed.
- *Travel* consists of mileage for assessments and observations and travel costs for the Principal Investigator to travel to Seattle at least one time per year.
- *Supplies* includes Picture Peabody Vocabulary Test IV and Woodcock-Johnson test books and scoresheets, CLASS scoresheets, ECERS-3 manuals, and other relevant supplies for observers and assessors. Costs increase in line with higher volumes of classrooms and children because more assessments, assessors, and observers are necessary.

Family Child Care (FCC) Pilot Evaluation Budget

The total budget for the Family Child Care Pilot Evaluation is $175,000 (see Table 13). The budget is composed of two parts: 1) development and 2) evaluation. The development, which will take place during the 2015–16 school year, is budgeted at $65,000. The evaluation is budgeted at $110,000; the timeline has not yet been determined.

Table 13: Four-year budget for FCC pilot evaluation

Adopted Budget: FCC Pilot Evaluation	SY 2015-16	SY 2016-17	SY 2017-18	SY 2018-19	Total
Pilot Evaluation Development	$65,000	$0	$0	$0	$65,000
Pilot Evaluation	$55,000	$55,000	$0	$0	$110,000
Total	**$120,000**	**$55,000**	**$0**	**$0**	**$175,000**

The development budget includes the following activities:

- *Limited Landscape Memo*: The Evaluation Team will submit a memo summarizing various approaches to offering high-quality preschool through an FCC setting.

- *Stakeholder Engagement Presentation:* DEEL will appoint members of the FCC Pilot Advisory Committee (with support from the Evaluation Team), after which the Evaluation Team will prepare a high-level presentation to engage this Advisory Committee.
- *FCC Pilot Evaluation Design Advisory Committee Meetings:* The Evaluation Team will facilitate the meetings to review the program elements of SPP and assess their applicability for FCC settings.
- *FCC Pilot Evaluation Plan:* The Evaluation Team will submit a report to DEEL summarizing the conclusions and recommendations of the Advisory Committee. Recommendations will focus on: 1) how FCC providers can participate in SPP and meet all program elements and 2) how FCC providers will be integrated into the Comprehensive Evaluation Strategy.

The development budget assumes that DEEL will be responsible for all logistics for the FCC Advisory Committee, including inviting individuals to join, securing spaces and acquiring food for its meetings, communicating with its members, and soliciting translators and child care, etc., as needed.

The activities associated with the evaluation budget have not yet been determined. The Evaluation Team will analyze options to work within the $110,000 budget when analyzing how FCC providers might be integrated into the Comprehensive Evaluation Strategy.

Deliverables

Please refer to Table 14 for a schedule of milestones and deliverables related to this budget.[81]

Table 14: Detailed deliverables with dates

Deliverable	Content	Date
Comprehensive Evaluation Strategy	Strategy document, literature review, PowerPoint presentation, and fact sheet.	August 2015
Process and Impact Evaluation Status Update	Short memo submitted separately for each evaluation type.	November 2015/16/17/18
FCC Pilot Evaluation Development	Landscape memo, stakeholder engagement presentation, facilitation of Advisory Committee, and FCC evaluation plan.	April 2016
Process Evaluation Mid-Year Report #1	Projected to include evaluation on professional development and curriculum implementation. This deliverable will also include expenses related to program management.	February 2016/17/18/19
Impact Evaluation Status Update	Short memo submission.	February 2016/17/18/19

[81] The Evaluation Team recommends that DEEL remunerate based on time and materials for work performed during the preceding time period at each milestone checkpoint. By doing so DEEL will align the work performed with the actual expenses incurred and thus avoid overpaying or underpaying the Evaluation Team. This engagement structure is more appropriate for evaluation than a deliverable-based contract in that it better accounts for the uncertainty naturally inherent in the flow of work.

Deliverable	Content	Date
Process Evaluation Mid-Year Report #2	Projected to include evaluation on classroom environments. This deliverable will also include expenses related to program management.	May 2016/17/18/19
Impact Evaluation Status Update	Short memo submission.	May 2016/17/18/19
Process Evaluation School Year (SY) 2015–16 Final Report	Projected to include evaluation program standards and system evaluation. This deliverable will also include expenses related to program management.	August 2016/17/18/19
Impact Evaluation SY2015–16 Final Report	All content related to impact evaluation.	August 2016/17/18/19
Consolidated SY2015–16 Executive Summary and Raw Data Transfer	Synthesized executive summary in one voice and a raw data transfer to DEEL.	September 2016/17/18/19
FCC Pilot Evaluation	Content to be determined based on the FCC Pilot Evaluation Assessment.	Date TBD

Appendices

Appendix A: Summary of Programs and Evaluation Designs used in the Technical Report for the City of Seattle.

Table A: Program Summary Description

Program	Enrollment	Eligibility	Teacher Qualification	Class Size	Hours	Curriculum	Professional Development	Annual Cost
Head Start	927,275	3, 4, & 5 year olds below FPL	30% with BA	4's: 20 3's: 17	3.5–6 hours/day 4–5 days/week	None prescribed	Varies by state	$7,726 plus 20% local match
Georgia	81,453	All 4's	BA	22	6.5 hours/day	Aligned w/ GA Early Learning Standards	15 hours/year	$2,238 state + local and federal
New Jersey Abbott	43,896	All 3's and 4's in Abbott districts	BA	15	6 hours/day	Aligned w/ NJ Preschool Teaching & Learning Standards	100 hours/5 years	$13,337
Oklahoma	40,823	All 4's	BA	20	2.5–6 hours/day	Must address PASS Standards for pre-K	75 hours/5years	$7,678
Rhode Island	234	Lottery	BA	18	6 hours/day	Aligned w/ RI Early Learning Standards	20 hours/year	$9,763
Michigan	30,552	<300% FPL; others pay fee	BA	18	3 or 6.5 hours/day	Aligned with MI Early Childhood Standards	6 credit hours/5 years	$5,704 plus local funding
North Carolina	26,617	At-risk 4's	BA	18	6.5 hours/day	Approved curriculum that aligns w/ NC Early Learning Standards	15 CEUs/5 years	$7,351
Tennessee	18,609	185% FPL	BA	20	5.5 hours/day	Varies	18 hours/year	$5,895

Program	Enrollment	Eligibility	Teacher Qualification	Class Size	Hours	Curriculum	Professional Development	Annual Cost
Cities								
Boston	2,300	Lottery	BA	22	6 hours/day	OWL literacy; Building Blocks math	Coach per 10 classrooms	$12,000
Denver	5,467	All Denver residents	Varies	Varies	Varies 3–8 hours/day	Varies	Varies	$419
San Antonio	1,500	Sliding scale	BA	20	7 hours/day	Frog Street Press and Teaching Strategies Curricula	14 professionals for collaboration with districts	$14,631
CPC	1,150 per cohort	Title I districts	BA	17	3 hours/day	Varies; language focused	In-service for administrators and teachers	$6,470

Table B. Evaluations Summary Description

State	Year	Design	n	Age Follow-up	Process Measure	Child Assessment	Outcomes					Evaluation Funding
							Math	Lang	Lit	EF	Soc/Em	
Head Start		RCT	2,559	3's end of pre-K	ECERS-R/FDCRS	WJ, PPVT, plus others Parent Report Teacher Report K-3	NS	+	+		+	Federal
			2,108	4's end of pre-K			NS	NS	+		NS	
				1st & 3rd grade			NS	NS	NS		NS	
Georgia and Oklahoma		DD	NAEP	8th grade	None	NAEP	+	NS				?
Georgia	2012	RDD	1,181	pre-K	CLASS ECERS-R	WJ, SSiS	+	+	+			State
	2011	DiD	>500,000	4th grade		NAEP	+	+				
Michigan	2004	RDD	865	pre-K			+	+	+			State
	2007	OLS	556	7th		MEAP test	+	+				
North Carolina	'03/09	RDD	992/1,010	pre-K	ECERS-R ELLCO CIS APEEC	PPVT TOPEL, WJ, other math	+	+	+			State
	'88/02	PSM	>200,000	3rd grade	ECERS-R	State test	+	+				
New Jersey	2004	RDD	2,075	pre-K	ECERS-R	PPVT, Lit WJ/WM NJASK	+	+	+			State and Foundation
			1038	K			+	+	+			
			1038/784	1st–5th grade			+	+				
Oklahoma	2004	RDD	836	pre-K		PPVT, Pre-CTOPPP, WJ	+	+	+	+		Federal and Foundation

State	Year	Design	n	Age Follow-up	Process Measure	Child Assessment	Outcomes					Evaluation Funding
							Math	Lang	Lit	EF	Soc/Em	
Rhode Island	2009	RCT	242	pre-K	ECERS-R CASEBA PRISM	PPVT, WJ, TOPEL, HTKS	+	NS	NS		NS	State and Foundation
Tennessee	2009	RCT	1,077	pre-K			+	+	+		NS	Federal
				K-2			NS	NS	NS			
Cities												
Boston	2008	RDD	2018	pre-K		PPVT, WJ , ERQ, TOQ, FDS, BDS, EF skills	+	+	+			Federal and City
Chicago	1986	PSM	1539	K, 12, 14, 24		ITBS soc/em	+	+				Federal and Foundation
	1986	PSM	1,150	K-8th grade			+		+			
Denver	2008	PSM	4,927	pre-K		TCAP, DRA	+		+			City
San Antonio	2013	OLS	555	pre-K	CLASS, Snap-shot	TSGOLD	+	NS	+		NS	City
Tulsa	2006	RDD	2,756	K	CLASS	WJ	+		+			Foundation

Appendix B: The 15 Essential Elements of High Quality Pre-K Programs

Excerpts from discussions with Jim Minervino

Lessons from Research and the Classroom[82] is rooted in a combination of:

- A comprehensive research review
- Interviews with the leaders of the exemplar programs highlighted
- Quantitative research undertaken by independent third parties (e.g., program evaluations done by NIEER in NJ, Christina Weiland, EdD, from Harvard for Boston), etc.

A meta-analysis was not used to determine the essential elements because a meta-analysis would result in the inclusion of many studies of programs that had no impact on children. Only four programs rigorously evaluated by third parties had meaningful outcomes on children that stuck with them for longer than a year or two.

Please refer to the Contributors for a list of experts consulted and contributors to the essential elements. Please also note that "Lessons from Research and the Classroom" was cowritten with Robert C. Pianta, PhD.

Summary of the 15 Essential Elements[83]

1. Education and compensation. Every lead teacher should have a B.A. plus an early learning credential and should earn compensation at the same level as K-3 teachers.
2. Adult-child ratios. Programs should have maximum class sizes of 22 children and adult-to-child ratios ranging from 2:15 to 2:22.
3. Learning time. Additional time in high-quality pre-K benefits children with the greatest needs. High-quality pre-K programs run 6 to 6.5 hours a day, 180 to 205 days a year.
4. Two adults in the classroom. High-quality pre-K programs have two adults in the classroom at all times: one lead teacher and one paraprofessional or aide.
5. Support for English language learners. Bilingual teachers and specialists can help students build skills in their native language while they learn English.
6. Support for students with special needs. For children with special needs, early intervention helps most. Early intervention can also reduce future special education needs and grade retention in elementary school.
7. Teacher-child interactions focused on learning. Teachers use structured activities and play to create rich learning environments in which children talk about what they are doing and follow their natural curiosity.
8. Age-appropriate learning standards. Goals for academic and social-emotional learning align with the expectations of Kindergarten and beyond.

[82] Minervino, Jim (with contribution from Robert C. Pianta, PhD, University of Virginia). Lessons from Research and the Classroom: Implementing High-Quality Pre-K that Makes a Difference for Young Children (White Paper, Bill & Melinda Gates Foundation, September 2014).
[83] Ibid.

9. Proven curriculum. Programs adhere closely to a research-based curriculum that is aligned with early learning standards and teachers' professional development.
10. Formative assessments. At the pre-K level, assessments are classroom-based and designed to help teachers and administrators improve outcomes for children.
11. Data-driven decision making. Programs use data to inform action and improve outcomes for children.
12. Professional development. Ongoing coaching focused on improving teacher-child interactions can help teachers improve instruction and student outcomes.
13. Integrated system. Standards, curriculum, professional development, formative assessments, and data are tied together and are mutually reinforcing.
14. Political will. Support from elected officials or a judicial mandate can help sustain public commitment to high-quality pre-K.
15. Strong leadership. Educators create a culture of high expectations, cultivate political will, and communicate the importance of quality instruction to parents.

Appendix C: More Information about CLASS, ECERS-3, and TS GOLD

Table C: Preschool CLASS dimension descriptions

Domain	Dimension	Description
Emotional Support	Positive Climate	Reflects the emotional connection between teachers and children and among children, and the warmth, respect, and enjoyment communicated by verbal and nonverbal interactions.
	Negative Climate	Reflects the overall level of expressed negativity in the classroom. The frequency, quality, and intensity of teacher and peer negativity are key to this dimension.
	Teacher Sensitivity	Encompasses the teacher's awareness of and responsiveness to students' academic and emotional needs.
	Regard for Student Perspectives	Captures the degree to which the teacher's interactions with students and classroom activities place an emphasis on students' interests, motivations, and points-of-view and encourage student responsibility and autonomy.
Classroom Organization	Behavior Management	Encompasses the teacher's ability to provide clear behavioral expectations and use effective methods to prevent and redirect misbehavior.
	Productivity	Considers how well the teacher manages instructional time and routines and provides activities for students so that they have the opportunity to be involved in learning activities.
	Instructional Learning Formats	Focuses on the ways in which teachers maximize students' interest, engagement, and abilities to learn from lessons and activities.
Instructional Support	Concept Development	Measures the teacher's use of instructional discussions and activities to promote students' higher-order thinking skills and cognition and the teacher's focus on understanding rather than on rote instruction.
	Quality of Feedback	Assesses the degree to which the teacher provides feedback that expands learning and understanding and encourages continued participation.
	Language Modeling	Captures the effectiveness and amount of teacher's use of language-stimulation and language-facilitation techniques.

Table D: ECERS-3 subscale descriptions

Subscale	Description
Space and Furnishings	This subscale addresses the areas of indoor and outdoor space, room arrangement, organization, display, furnishings and equipment.
Personal Care Routines	This subscale addresses practices around daily routines like greeting and departure, meals, naptime, and toileting as well as health and safety practices.

Subscale	Description
Language-Reasoning	This area addresses the classroom's formal and informal communication, language, and reasoning opportunities.
Learning Activities	This subscale looks at the learning opportunities in each of the areas of the classroom including fine motor, art, music/movement, blocks, sand/water, dramatic play, nature/science, math/number, use of video/computer, and diversity.
Interactions	This area addresses supervision of children, discipline, staff-child interactions, and interactions among children.
Program Structure	This area addresses classroom operations and schedule, including groupings, transitions, and flexibility.

Table E: TS GOLD student assessment description

Areas of Development and Learning	Objectives
Social–Emotional	1. Regulates own emotions and behaviors 2. Establishes and sustains positive relationships 3. Participates cooperatively and constructively in group situations
Physical	4. Demonstrates traveling skills 5. Demonstrates balancing skills 6. Demonstrates gross-motor manipulative skills 7. Demonstrates fine-motor strength and coordination
Language	8. Listens to and understands increasingly complex language 9. Uses language to express thoughts and needs 10. Uses appropriate conversational and other communication skills
Cognitive	11. Demonstrates positive approaches to learning 12. Remembers and connects experiences 13. Uses classification skills 14. Uses symbols and images to represent something not present
Literacy	15. Demonstrates phonological awareness 16. Demonstrates knowledge of the alphabet 17. Demonstrates knowledge of print and its uses 18. Comprehends and responds to books and other texts 19. Demonstrates emergent writing skills
Mathematics	20. Uses number concepts and operations 21. Explores and describes spatial relationships and shapes 22. Compares and measures 23. Demonstrates knowledge of patterns

Appendix D: Potential Assessment Additions

We recommend including the following instruments as part of the evaluation in order to better measure children's development across a wider range of dimensions, as well as to better measure the supports in the classroom for dual language acquisition and for the learning of literacy, math, and science, among other skills.

Child assessments

The use of any of the following assessments would provide a more encompassing picture of children's progress. We provide alternatives that allow gaining greater depth in a particular dimension or aspect of a child's development or additional dimensions not taken into account in the battery proposed above. Even if this information is captured for a subsample of programs and classrooms, it would complement the existing battery well.

1. *The Early Literacy Skills Assessment* (*ELSA)* (Debruin-Parecki, 2005) would strengthen the existing battery by measuring a broad range of language and literacy constructs, including comprehension, phonological awareness, alphabetic principle, and concepts about print. This measure is available in Spanish and English with norms in both.
2. The *Early Writing Assessment* (*EWA)* (Puranik & Lonigan, 2011) is a measure of written language knowledge available in English and Spanish (although Spanish is unnormed) that captures the linear sequence in writing skills acquisition in preschoolers.
3. The *Lens on Science (LENS),*[84] available in English only, is a comprehensive science measure that, although expensive to administer, allows capturing in detail children's development in this content area.
4. Similarly, The *Tools for Early Assessment in Math* (*TEAM)*, short-form adaptation of the REMA, (Weiland, et al., 2012) assesses a wide range of early numeracy, geometry, and spatial skills. It was recently used in the Boston evaluation.
5. The *Child Behavior Checklist (CBCL)*[85] has a teacher-rated, low-cost form (C-TRF) that could be used in the pre-post component to have a measure of children's socio-emotional development. Available in English and Spanish, this measure would serve as a complement to the proposed battery.

Classroom assessments

The use of any one of the following measures would strengthen understanding of various aspects of quality that matter for children's development. It is important to take into account that commonly used measures such as CLASS and ECERS have only explained on average about 10 percent of the variation observed in children, and that, while very useful for understanding quality in the program and strengthening it over time, no one measure is really the ultimate indicator of quality. As a consequence, the more dimensions and aspects of quality that can be captured, even if this is done in a subsample of classrooms, the more depth that will be gained in understanding the processes and supports observed inside the classrooms and which of these matters for children's development in the SPP.

A measure of the use of time in classrooms such as the *Snapshot* (Ritchie, Howes, Kraft-Sayre, & Weiser, 2002) can be particularly important for understanding what is and is not present in classrooms that would

[84] http://ies.ed.gov/ncer/projects/grant.asp?ProgID=7&grantid=805
[85] http://www.aseba.org/preschool.html

explain any particular trends or deficits observed in children's development. This observation tool measures how children and teachers spend their time in the classroom. Used in conjunction with global measures of classroom quality in national studies, the Snapshot has been shown to predict child progress. The Snapshot has good interobserver reliability, with a kappa value of .95 (Pianta, Howes, Burchinal, Bryant, Clifford, Early, and Barbarin, 2005). Observations consist of time-sampled codes assigned to teacher and child behaviors, every 60 seconds (representing one cycle) over the course of the morning. Typically, four children are randomly selected from each classroom; and each child is observed for 40 seconds, followed by 20 seconds of coding. This sequence is repeated for 2 to 3 hours in each classroom. Codes are divided into five subscales, including activity setting (i.e., whole group, free choice, transitions); peer interaction (simple social, cooperative pretend); child engagement (i.e., science, mathematics, oral language development); teacher-child engagement (i.e., scaffolds, didactic); and one-on-one teacher-child interactions (elaborated, routine).

A dual language supports measure such as *The Classroom Assessment of Supports for Emergent Bilingual Acquisition* (*CASEBA)* (Freedson, Figueras-Daniel, & Frede, 2008) would address the cultural concerns that have been raised for ECERS and, more particularly, for CLASS. The CASEBA is designed to assess the degree to which preschool teachers and classrooms provide support for the social, cognitive, and linguistic development of preschool-aged DLLs, with a focus on language and literacy. The instrument consists of 26 distinct rating scale items that cluster around six broad aspects of the early childhood curriculum: 1) teacher knowledge of child background information, 2) supports for home language and literacy development, 3) supports for English language and literacy development, 4) social-emotional supports and classroom management, 5) curriculum content, and 6) assessment. Each item is rated on a seven-point Likert scale, where 7 indicates that a specific form of support and accompanying practices are present in close to an ideal form, while 1 represents the total absence of any such practices. A modified version of this instrument can also be used to assess language and literacy teaching practices and supports for all children regardless of home language. Although the CASEBA has not yet been publicly disseminated, a validity study of the measure was conducted in 100 classrooms in New Jersey as well as in a staffing study of DLLs in an urban district in New Jersey. Findings from the validity study can be found in *Dual Language Learners in the Early Childhood Classroom* (Freedson, Figueras-Daniel, Frede, Jung, & Sideris, 2011). Although the tool is unpublished, it has garnered attention from early childhood dual language researchers across the country who are seeking observational tools to use in classrooms dominated by DLLs to assess the quality of teacher input and interactions for both research and professional development. Most recently CASEBA was highlighted as a valuable tool for focusing on the extent to which teacher practices and classroom quality specifically address the needs of DLL children (Castro, Espinosa, & Paez, 2011).

With a focus on literacy, *The Early Language and Literacy Classroom Observation* (*ELLCO)* (Smith, Brady, & Anastasopoulos, 2008) is a comprehensive set of observation tools for measuring the level of classrooms supports for language and literacy development. The ELLCO pre-K includes five sections: classroom structure, curriculum, the language environment, books and book reading, and print and early writing. These five sections are grouped into two main subscales: the general classroom environment subscale, which consists of the classroom structure and curriculum sections, and the language and literacy subscale, which includes the language environment, books and book reading, and print and early writing sections. Reliability analyses have shown high Cronbach's alphas ranging from .723 for the curriculum section to .894 for the print and early writing section. The tool has also shown good test–retest reliability and sensitivity to interventions that target literacy.

A measure of math and science supports such as *The Preschool Rating Instrument for Science and Mathematics* (*PRISM)* (Stevenson-Boyd, Brenneman, Frede, & Weber, 2009) would capture more detailed evidence of the types and quality of supports for math and science that are available in each classroom relative to the ECERS-3. The PRISM is a 16-item classroom observation instrument designed to measure the quality of materials and staff interactions to support preschoolers' mathematics and science learning. The 11 mathematics items in the PRISM provide a comprehensive picture of instructional supports for a wide range of mathematics skills and reasoning, including typically neglected areas such spatial reasoning, patterns, sequencing, and measurement; and its items and indicators are informed by the NAEYC/NCTM (2002). Two science items on the PRISM focus on materials that support explorations of biological and nonbiological science and encourage reading and writing about science. Three items focus on teaching interactions that encourage children to investigate, experiment, and discuss scientific concepts; support observing and predicting; and encourage children to record science information from their investigations. All PRISM items are scored on a seven-point scale with anchor points at the odd numbers: 1 = no materials or interaction or poor quality; 3 = minimal quality; 5 = good quality; and 7 = excellent quality. At each level, one or more indicators serve as evidence that a classroom has achieved that level of quality. The PRISM is currently being used in large-scale studies in New Mexico and New Jersey.

References for potential additions

Castro, D. C., Espinosa, L., & Páez, M. (2011). Defining and measuring quality early childhood practices that promote dual language learners' development and learning. Quality measurement in early childhood settings, 257–280.

DeBruin-Parecki, A. (2005). Early literacy skills assessment. High/Scope Educational Research Foundation.

Diamond, A., & Taylor, C. (1996). Development of an aspect of executive control: Development of the abilities to remember what I said to "do as I say, not as I do." *Developmental Psychobiology, 29*, 315–344.

Dunn, D. M., & Dunn, L. M. (2007). Peabody picture vocabulary test: Manual. Pearson.

Dunn, L. M., Lugo, D. E., Padilla, E. R., & Dunn, L. M. (1986). Test de Vocabulario en Imagines Peabody. Circle Pines, MN: American Guidance Service.

Early, D. M., Maxwell, K. L., Burchinal, M., Alva, S., Bender, R. H., Bryant, D., Cai, Karen, Clifford, R., Ebanks, C., Griffin, J., Henry G., Howes, C., Iriondo-Perez, J., Jeon, H., Mashburn, A., Peisner-Feinberg, E., Pianta, R., Vandergrift, N., & Zill, N. (2007). Teachers' education, classroom quality, and young children's academic skills: Results from seven studies of preschool programs. *Child Development*, *78*(2), 558–580.

Feuer, M. J., Towne, L., & Shavelson, R. J. (2002). Scientific culture and educational research. *Educational Researcher, 31*(8), 4–14.

Freedson, M., Figueras-Daniel, A., & Frede, E. Classroom Assessment of Supports for Emergent Bilingual Acquisition (2008). New Brunswick, NJ: National Institute for Early Education Research, Rutgers, The State University of New Jersey.

Freedson, M., Figueras-Daniel, A., Frede, E., Jung, K., & Sideris, J. (2011). The Classroom Assessment of Supports for Emergent Bilingual Acquisition: Psychometric properties and key initial findings from New Jersey's Abbott Preschool Program. In C. Howes, J. T. Downer, & R. C. Pianta (Eds.), Dual Language Learners in the early childhood classroom, 233–258, Baltimore, MD: Paul H. Brookes.

Frye, D., Zelazo, P. D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive Development*, *10*(4), 483–527.

Harms, T., Clifford, R. M., & Cryer, D. (2014). Early childhood environment rating scale. Teachers College Press.

La Paro, K. M., Pianta, R. C., & Stuhlman, M. (2004). The classroom assessment scoring system: Findings from the prekindergarten year. *The Elementary School Journal*, 409–426.

Meador, D. N., Turner, K. A., Lipsey, M. W., & Farran, D. C. (2013). Administering measures from the PRI learning-related cognitive self-regulation study. Nashville, TN: Vanderbilt University, Peabody Research Institute.

Muñoz-Sandoval, A. F., Woodcock, R. W., McGrew, K. S., Mather, N., & Schrank, F. A. (2005). Batería III Woodcock-Muñoz: Suplemento diagnóstico para las pruebas de habilidades cognitivas. Itasca, IL: Riverside Publishing.

Osborn, A. F., Butler, N. R., & Morris, A. C. (1984). The Social Life of Britain's Five-year-olds (Henley, Routledge & Kegan Paul).

Pianta, R., Howes, C., Burchinal, M., Bryant, D., Clifford, R., Early, D., & Barbarin, O. (2005). Features of pre-kindergarten programs, classrooms, and teachers: Do they predict observed classroom quality and child-teacher interactions? *Applied Developmental Science*, *9*(3), 144–159.

Pianta, R. C., LaParo, K. M. & Hamre, B. K. (2006). Classroom Assessment Scoring System: Manual Pre-K. Charlottesville, VA: Center for Advanced Study of Teaching and Learning.

Pianta, R. C., LaParo, K. M., & Hamre, B. K. (2008). Classroom Assessment Scoring System (CLASS) Manual, Pre-K. Baltimore, MD: Paul H. Brookes Publishing.

Ponitz, C. E. C., McClelland, M. M., Jewkes, A. M., Connor, C. M., Farris, C. L., & Morrison, F. J. (2008). Touch your toes! Developing a direct measure of behavioral regulation in early childhood. *Early Childhood Research Quarterly, 23*(2), 141–158.

Puranik, C. S., & Lonigan, C. J. (2011). From scribbles to scrabble: Preschool children's developing knowledge of written language. *Reading and Writing, 24*(5), 567–589.

Ritchie, S., Howes, C., Kraft-Sayre, M., & Weiser, B. (2002). Emerging academics snapshot. Los Angeles: University of California.

Smith, M. W., Brady, J. P., & Anastasopoulos, L. (2008). Early Language and Literacy Classroom Observation Tool: Pre-K. Paul H. Brookes Publishing Company.

Stevenson-Boyd, J. S., Brenneman, K., Frede, E., & Weber, M. (2009). Preschool rating instrument for science and mathematics. New Brunswick, NJ: National Institute for Early Education Research, Rutgers, The State University of New Jersey.

Weiland, C., Wolfe, C. B, Hurwitz, M. D., Clements, D. H., Sarama, J. H. & Yoshikawa, H. (2012): Early mathematics assessment: validation of the short form of a prekindergarten and kindergarten mathematics measure, *Educational Psychology: An International Journal of Experimental Educational Psychology, 32*(3), 311–333.

Wong, V. C., Cook, T. D. Barnett, W. S., & Jung, K. (2008). An effectiveness-based evaluation of five state pre-kindergarten programs. *Journal of Policy Analysis and Management, 27*(1), 122–154.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). Woodcock-Johnson III Tests of Achievement. Rolling Meadows, IL: The Riverside Publishing Company.

Zelazo, P. D. (2006). The Dimensional Change Card Sort (DCCS): A method of assessing executive function in children. http://www.nature.com/search?journal=nprot&q=dimensional%20change%20card%20sort