Summer Learning Program Quality Intervention (SLPQI)

Phase Two Feasibility Study

Charles Smith, Ph.D.

Ravi Ramaswamy

Anna Gersh, Ph.D.

Gina McGovern

Executive Summary

In 2013, the David P. Weikart Center (Weikart Center) and National Summer Learning Association (NSLA) began a collaboration to improve the quality of summer learning programs. The *Summer Learning Program Quality Intervention* (SLPQI) is a continuous improvement intervention for summer learning programs that includes four core components: (1) a standard and measures for quality of management and instructional practices – and the *Summer Learning Program Quality Assessment* (Summer Learning PQA); (2) training and technical assistance supports, (3) performance data products and (4) a continuous improvement cycle that fits the prior three elements to local circumstances and resources. The SLPQI and the Summer Learning PQA were designed to advance the science and practice of continuous improvement by focusing on qualities of learner experience that optimize skill building in specific domains, *active-participatory* and *academic*.

This report provides findings for an evaluation of the implementation of the SLPQI in 32 summer learning programs in Grand Rapids, Mich., Northern California, and Seattle, Wash. The evaluation addressed feasibility of implementation, customer satisfaction and the effectiveness of the Summer Learning PQA to differentiate higher from lower quality programs. Key findings include: (1) overall satisfaction with SLPQI and Summer Learning PQA was high; (2) implementation of SLPQI was uneven across sites, although partially by design; (3) readiness for SLPQI can be characterized by several key features and timing is the greatest challenge; (4) precision and meaningfulness of Summer Learning PQA data is promising; (5) summer learning programs contained a mix of higher and lower quality features, with differing profiles for active-participatory versus academic instructional quality.

Organizational Background

In 2013, the David P. Weikart Center (Weikart Center) and National Summer Learning Association (NSLA) began a collaboration to address summer learning program quality improvement.

The Weikart Center's mission is to empower education and human service leaders to adapt, implement and scale best-in-class, research-validated quality improvement systems to advance child and youth development. The Weikart Center currently provides systems-level consulting, coaching, technical assistance and training anchored by the Program Quality Intervention (PQI) and/or a Program Quality Assessment (PQA) brands. In 2014, over 4,000 out-of-school time programs in 105 systems and 38 states employed the PQI and/or a PQA. These systems support both continuous improvement and evaluation of system-level performance, cost effectively and at scale. The Weikart Center is an affiliate division of the Forum for Youth Investment.

The National Summer Learning Association (NSLA) is the only national nonprofit exclusively focused on closing the achievement gap by increasing access to high quality summer learning opportunities. As the leading national expert on summer learning program quality and community systems building, NSLA works with individual and groups of summer programs and community steering committees to use data to reflect on the state of their program or system, develop and implement a quality improvement plan and assess the results of their efforts against research-based indicators of quality. NSLA recognizes and disseminates what works, offers expertise and support for programs and communities and advocates for summer learning as a solution for equity and excellence in education.

Table of Contents

I. Introduction and Overview of the Study

Summer learning programs are positioned to play an important role in reducing summer learning losses that disproportionately affect disadvantaged students (Alexander, Entwisle, & Olson, 2007; Harris Cooper, Nye, Charlton, Lindsay, & Greathouse, 1996) and summer learning programs with an explicit focus on improving academic skills are an important part of the out-of-school time landscape (Boss & Railsback, 2002; Newhouse, Neely, Freese, Lo, & Saili, n.d.). While a growing literature suggests that summer learning programs can impact academic and other school-related skills (Borman & Dowling, 2006; Chaplin & Capizzano, 2006; McCombs, Augustine, & Schwartz, 2011; McCombs et al., 2014; Roderick, Engel, & Nagaoka, 2003), few rigorous studies have closely examined the specific features and practices that mediate or moderate relationships between summer program participation and school success outcomes (Arbreton et al., 2008; Spielberger & Halpern, 2002).

This relatively oblique understanding about the specific practices that support skill development in young learners (learners with unique profiles of individual assets and locally defined individual needs, i.e., skills necessary to be successful in my school and my classroom next year) presents a number of challenges to the potential of summer learning programs. First, without sufficiently detailed descriptions of practice, it is impossible to evaluate the effectiveness of specific practices as an extension from summative impact evaluation designs.[1] Second, and perhaps more importantly, without measures of practice that are both sufficiently precise and feasible to implement, it is difficult to provide either standards that drive planning for high quality services or performance feedback necessary for accountability and service improvement.

The *Summer Learning Program Quality Intervention* (SLPQI) and the *Summer Learning Program Quality Assessment* (Summer Learning PQA) directly address these challenges. The SLPQI is a continuous improvement intervention for summer learning programs that includes four core components: (1) a standard and measures for quality of management and instructional practices – the Summer Learning PQA, (2) training and technical assistance supports, (3) performance data products and (4) a continuous improvement cycle that fits the prior three elements to local circumstances and resources. The SLPQI and

---

[1] The fields of education and human services are currently tuned to hear results from outcome studies, over-valuing impact evaluation and its methodological emphases (e.g., average effects, reliability of measures to detect individual differences, internal and external validity of research design, statistical power) and undervaluing the science and practice of continuous improvement (Hiebert, Gallimore, & Stigler, 2002) and its methodological emphases (individual growth and change, moderation/mediation, implementation, criterion validity, feasibility). As Larson (2000) aptly noted over a decade ago, "Outcomes research is often the necessary evil that is done before anyone knows what to look for…. Evaluative data have been needed to justify funding for youth activities, even though we are not yet sure what the independent and dependent variables for this evaluative research should be."

the Summer Learning PQA were designed to advance the science and practice of continuous improvement by focusing on qualities of learner experience that optimize skill building in specific domains, *active-participatory* and *academic*.

Because explicit educational curricula have proven difficult to implement with fidelity or at scale, the identification of best practices related to student learning has been of increasing interest (Jones & Bouffard, 2012) and the continuous improvement approach may prove to be an effective way to bring best practices to scale (Smith & Akiva, 2008). The SLPQI and Summer Learning PQA represent an opportunity to generate sustained conversations among expert practitioners in thousands of programs, with performance data in hand, about what the standard for high quality summer learning services should be and what kinds of program designs are necessary to achieve that high standard. Because the sequence of design, iteration, and validation for scalable social innovation is both long and expensive, seeding many of these local conversations may produce greater near-term social value – in terms of identification of best practices and innovation in program design – than concentrated investments in impact trials for under-developed and under-implemented curricula.

Overview of the Study

An earlier Phase I proof-of-concept pilot was conducted to initiate development of the Summer Learning PQA standard/measure[2] and explore application of this standard/measure in a continuous improvement cycle at 16 sites in Baltimore, Grand Rapids and Oakland. In this earlier phase, Weikart Center and NSLA also began to lay the foundations for long term collaboration around a joint product and client base. A key finding from Phase I was that program staff described the continuous improvement cycle as useful and "worth their time." In particular, the Phase I pilot identified opportunities to better fit the continuous improvement cycle and the PQA measures to the context of summer learning programs. Additional detail is available in a report (Ramaswamy, Gersh, Sniegowski, McGovern, & Smith, 2014).

The Phase II study was designed to test the feasibility of implementing a continuous improvement sequence at scale in place-based summer learning systems. We also hoped to advance our understanding of the reliability and validity of the information produced by the Summer Learning PQA measures. In the Phase II proposal the following goals were listed:

- Further clarification of the quality standard at the organization and instructional levels

---

[2] Note on language: When considered together, the level-5 descriptors in the Program Quality Assessment rubrics are a quality standard (singular) for management and instruction in the out-of-school time field. In this sense, the performance standard and performance measure are integrated.

- Improvement of quality measures and provision of detail on reliability and validity to support both performance improvement and program evaluation uses
- Refinement of the SLPQI and delivery in approximately 40 sites with varied designs (e.g., school and CBO; teacher and volunteer led; systems with existing quality improvement infrastructure and systems without; school-age and youth).

While the Phase II study dedicated the largest proportion of resources to design, development and implementation, we also conducted an evaluation of the implementation portion of the project. The evaluation addressed (1) feasibility of implementing the SLPQI, (2) customer satisfaction and (3) the effectiveness of the Summer Learning PQA to differentiate higher from lower quality programs.

The project's aggressive eight month timeline extended from February through September of 2014. Part II of this report describes the four elements of the SLPQI as implemented during the study. Part III describes the evaluation sample and procedures. Part IV describes fidelity of implementation of the SLPQI at 32 summer learning program sites in four cities as well as overall customer satisfaction. Part V provides an aggregate performance profile for the 32 sites. Finally, in part VI we discuss findings and recommendations for next iterations of the work.

Several appendices are also included. Appendix A presents descriptive statistics for each item, scale and composite score in the Summer Learning PQA Forms A (instructional practices) and B (management practices). Appendix B provides a logic model for summer programs fit to the SLPQI measures, followed by an abridged review of literature for each of the domains in the Summer Learning PQA. Appendix C presents technical detail regarding the reliability and validity of measures. Appendix D provides detail regarding missing data.

II. Summer Learning Program Quality Intervention (SLPQI)

Summer learning programs have unique characteristics to which a continuous quality improvement sequence must be fit. First, summer learning programs often combine grade-level instruction in explicit academic content with positive youth development experiences so it is critical to have a standard for management and instructional practices which are appropriate for delivery of this mix of content, over either a full-day or part-day. Second, programs are often of short duration (4-12 weeks) requiring implementation of a very rapid cycle[3] of continuous improvement, and therefore an organizational culture of sufficient capacity.

The elements and sequence of the SLPQI, as implemented in this study, are summarized in Figure 1 and described in greater detail in the remainder of this section. The SLPQI is modeled after the Youth Program Quality Intervention (YPQI), an evidence-based continuous improvement model for out-of-school time programs.[4] The SLPQI represents an adaptation of the YPQI to suit summer program content, structure and timelines.

Figure 1 summarizes the SLPQI sequence. In general, system leads receive technical assistance designing an organization-level continuous quality improvement cycle that they can both expect programs to implement and for which they can provide adequate supports across sites. Then, program managers and staff are engaged with workshops and training to learn the methods entailed by continuous improvement cycle and plan for implementation. Next, an assessor visits each site and produces a performance report based on interview and observation data and this report is passed back to program managers. Program managers then use the performance feedback during individual interactions with instructional staff as well as during improvement planning workshops that can occur during the summer program session, after the session in preparation for the following school year and summer, or both.

Four SLPQI elements – quality standard/measure, supports, data products, and improvement cycle – are each described below. Our experience suggests the following design principle for continuous quality improvement systems: the specific configuration of the organization-level continuous

---

[3] While rapid cycles may be advantageous (Cohen-Vogel et al., 2014), they are difficult to achieve. The summer learning programs and system we encountered might benefit more from a longer arc of repeated cycles over multiple years.

[4] The Youth Program Quality Intervention is the most widely used quality assurance process in the afterschool field and was the subject of a randomized trial which demonstrated that high fidelity to the same four continuous improvement elements improved the quality of instructional experiences for at-risk youth (Smith, Akiva, et al., 2012). Subsequent validation studies have linked exposure to high quality instructional practices, as defined by the Youth Program Quality Assessment, to improved school success outcomes (Naftzger, 2014; Naftzger et al., 2013; Naftzger, Tanyu, & Stonehill, 2010; Naftzger, Vinson, Manzeske, & Gibbs, 2011).

improvement sequence and the system-level supports will likely vary considerably across organizations. For example, some organizations may be prepared to implement immediately and have many system-level supports in place, whereas other organizations may need several months of preparation before implementation and have few system-level supports in place. The essential task then is to provide a "design standard"[5] for the core elements so that system leaders may innovate toward deeper and more successful implementation. This report provides a preliminary design standard for the SLPQI: a discussion of which parts of the SLPQI are the key elements, what the results of prior implementations reveal about successes and pitfalls, and how the key elements can fit together in optimal ways given different resources and constraints. In terms of the science of evaluation, a design standard supports clarity around the intervention elements which must be adopted outright and those elements which can be adapted to fit local circumstances (Smith, 2013).

Figure 1

SLPQI Theory of Action

---
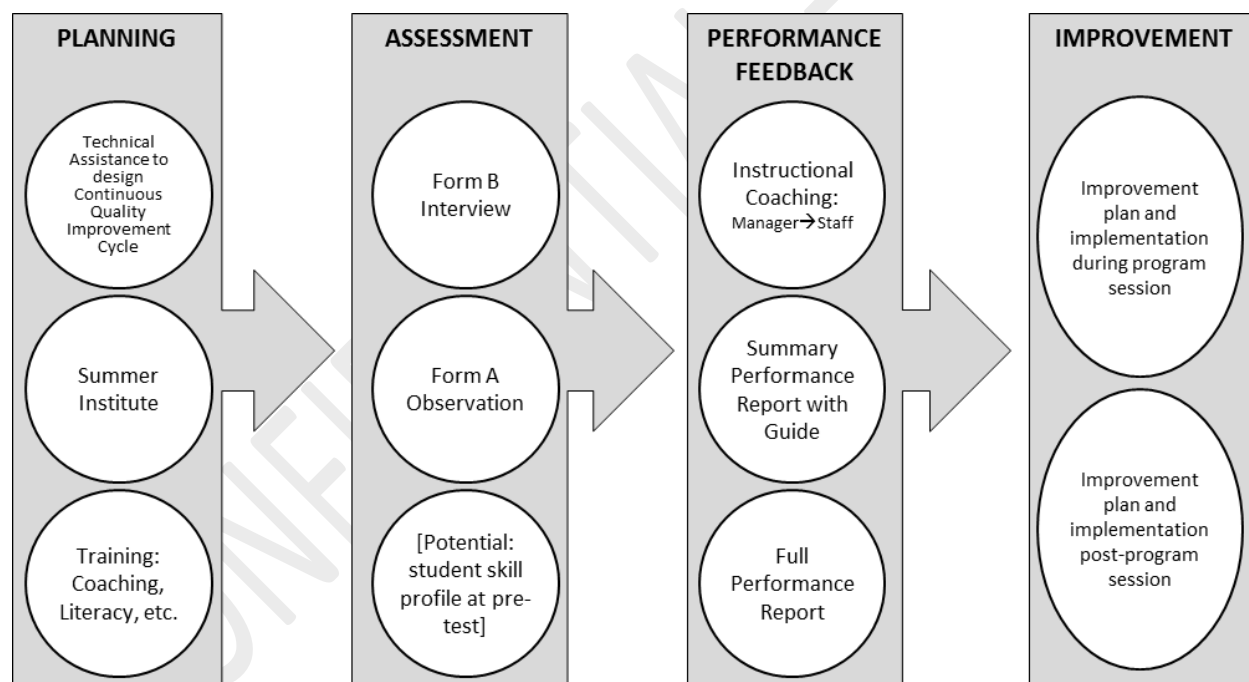
[5] We use the term "design standard" to mean something like product requirements in private sector usage – which means "what you have to do to make a product that satisfies specific customer demands or values." See discussion in Smith (2013).

Standard/Measure: Summer Learning Program Quality Assessment (Summer Learning PQA)

In the Phase II study, the Summer Learning PQA Forms A and B were revised to include both new scales and several revised items. Each site was visited by a trained external assessor who collected observation and interview data during a one-day site visit. Table 1 describes the method used to complete each form and the measurement objective achieved.

Table 1

Component Quality Measures of the Summer Learning PQA

Measure	Method	Objective
Form A Program Observation	Two 90-minute periods of program observation spanning various program offerings; one AM and one PM	To rate the quality of specific instructional practices and produce an overall quality rating for the site
Form B Program Manager Interview	30-minute interview with a designated site coordinator	To rate the quality of specific management practices

The data collection method for the Form A observation was configured to produce quality ratings that best represent the most prevalent summer program designs: academic content in the morning session and enrichment in the afternoon – or academically-focused morning-only programs more like traditional summer school.[6] The amount and timing of data collection during the one-day assessor visit was designed to optimize meaningfulness of sub and composite scores across both levels of measurement (e.g., item, scale, domain, total) and levels of setting (e.g., classrooms within program within system).[7]

The Form B interview with the program manager assesses management practices in four domains, largely derived from NSLA's Comprehensive Assessment of Summer Programs (Kreider, Caspe, Kennedy, & Weiss). The full set of Summer Learning PQA measures are provided in in Appendix A. An abridged review of the pertinent literature for each domain of Summer Learning PQA is provided in

---

[6] A typical *summer school* program usually only provides remediation in the student's previous level of learning in order to advance to the next grade. A typical *summer enrichment* program (like a summer camp) usually provides no academic instruction, but rather an entire day of enrichment activities and field trips that are not always intentionally linked to learning. The ideal summer program as described by youth and parents is different from the regular school day (Duffett, Johnson, Farkas, Kung, & Ott, 2004).

[7] A literature exists on the number of observations necessary to represent instruction delivered across occasions (Bill and Melinda Gates Foundation, 2010; Smith, Akiva, et al., 2012; Whitehurst, Chingos, & Lindquist, 2014). Often framed as a reliability issue, the key point is that multiple observations are necessary to represent teacher practice with a high degree of precision. The Weikart Center has also examined the issue of reliability of program scores that combine multiple observations across different staff into a single program rating and, that is the logic of this research that we have followed in developing the method to generate quality ratings for summer programs (Smith, 2013).

Appendix B. Appendix C provides a detailed discussion of reliability and validity for the revised Form A measure.

Supports: Training and Technical Assistance

Several types of supports are provided through the SLPQI, including (1) technical assistance for system leaders and quality intermediary organizations (QIOs) supporting summer learning systems, (2) training for program managers on the content of the SLPQI and (3) training for program assessors.

Technical assistance for system leaders and QIOs includes ongoing contact with an experienced project manager who can provide guidance and recommendations for timelines, recruitment and creative problem-solving in real time. System leaders also have access to quality improvement system artifacts from other networks across the country and multiple ongoing opportunities to join learning communities focused on quality improvement system-building.

SLPQI trainings for program managers include the Summer Learning Institute, Instructional Coaching and Planning with Data.

Summer Learning Institute. The NSLA Summer Learning Institute kicks off the summer program cycle and is designed to provide an introduction to research around summer learning quality and promising practices anchored in NSLA's *Summer Starts in September* guidebook as well as management practices and tools that support program planning. In this project, the SLPQI process was also reviewed during the Summer Learning Institute. As a part of the Campaign for Grade Level Reading, Northern California programs received an additional workshop on the importance of early childhood literacy in eliminating the achievement gap, including specific tools and activities for improving literacy instruction. The literacy scale in the Summer Learning PQA was aligned with this content.

Instructional Coaching. This one-day training for site coordinators provides participants with skills necessary to observe staff in the process of instruction and to provide strengths-based feedback to teachers using the "observation-feedback" method (Akiva, 2007) aligned with the specific items from the Summer Learning PQA Form A. This improvement strategy can be implemented by a program manager at any time during the program session.

Planning with Data. This one-day workshop was adapted to support continuous improvement planning with the SLPQI performance report. During the workshop, program managers are supported to interpret SLPQI performance data and develop improvement plans targeting priority areas that they select. Rather than focusing on improvement planning only at the end of the session, in the Phase II project, we also asked program managers to implement plan for improvement during the summer session based on their performance report, but this proved difficult. In several cases, the improvement planning effort came after the end of the program and was focused on the following year.

Training and technical assistance is also available for assessors who conduct site visits and produce the performance reports for sites. The core training for assessors is the two-and-a-half-day *External Assessor Reliability Training* which culminates in a test for endorsement as a reliable external rater. For assessors who have received the endorsement in prior years, a half-day refresher webinar is available, as well as assessor check-in tests to address issues of rater drift during periods of intensive data collection.

Technical assistance for assessors and those coordinating the data collection process include co-development of data collection design and protocol, templates and best practices for scheduling assessments, accuracy checks and providing expert responses to scoring questions.

Three tools for assessors were also developed to support assessor compliance with the data collection method. First, a complete data collection protocol was developed to support preparation for the assessment visit and as a reference during the observation. Second, assessors received a hard copy or electronic version of an observation notebook for each site visit. Finally, an Excel-based workbook was created to increase automation in production of the performance report. The workbooks were submitted online and the reports were processed by the project team – with less than two-day turnaround in most cases. The performance reports will be further automated with expansion of the Weikart Center's Online Scores Reporter software to accommodate the Summer Learning PQA in 2016.

Performance Data Products

Following the assessor visit, program managers received a performance report – a package of performance data and tools for interpretation and use – describing management and instructional practices. In addition to item and scale level scores for all SLPQI measures, the report also included:

- One-page guide about how and where the data could be used during their summer session
- One-page overview of the quality standards referenced by the performance data
- Take-it-back agenda for a 30-minute workshop on the Summary Report
- Frequently asked questions on how to interpret the Summary Report data
- One-page narrative summary of strengths, suggested improvement actions, and other specific feedback from the assessor

In addition to the performance report for a specific program, an aggregate report on quality management and instructional practices across programs in a system was also created as part V of this report.

Improvement Cycle

The SLPQI improvement cycle occurs within each distinct program (i.e., at the organization level) but requires planning across programs (i.e., at the system level). The cycle is essentially a set of

dates describing when parts of the SLPQI get implemented and plans to receive support for implementation. Determining the sequence of elements that support the program level improvement cycle is a critical part of the technical assistance that system leaders receive early in the process. Furthermore, we encourage system leads to think in terms of multiple cycles (i.e., multiple years), over which capacity is incrementally developed. While essentially a set of dates and plans, the improvement cycle can provide a foundation for the continuous improvement culture that is the ultimate organization level objective for the SLPQI. If the improvement cycle is not carefully designed – and therefore implementable – it will not produce an effect on how the program managers and staff do their work.

III. Study Sample and Procedures

The evaluation component of the Phase II Feasibility Study addresses the following questions: How feasible was implementation of the SLPQI? How can the feasibility be improved? Was the SLPQI useful and therefore a good use of time? Did the Summer Learning PQA identify areas of low quality?

Sample

In the late winter and early spring of 2014, the project team recruited several summer learning systems to participate in the Phase II study. The study sample included both returning program sites from the earlier Phase I project as well as new sites. While over 40 programs were initially recruited into the study sample, a total of 32 summer learning programs participated in the study by achieving the minimum level of participation: completion of the Summer Learning PQA assessment.[8] Table 2 provides detail regarding the study sample.

Table 2
Participating Phase II System Characteristics

System in Phase II Cohort	Sites	Grades Served	Average Daily Attendance	Fee Based?	Ave Program Duration (weeks)
Northern California*	12	Pre-K – 8th Grade	120	5 Fee-based	5.3
City of Seattle	9	8th-11th (one K-5 site)	62	No fees	4.6
School's Out Wash – Feed Your Brain	4	K-6th (one K-12 site)	64	No Fees	4.4
Seattle Public Schools **	4	Elementary	100	No Fees	4.0
Grand Rapids	3	1st – 8th	30	No fees	4.3

*Includes Oakland site. ** Seattle Public Schools includes 17 programs organized into 4 physical locations, or sites. By the end, 3 of these programs had either dropped out or provided insufficient data to be included in the final analyses.*

---

[8] Sample sizes can be difficult to determine in evaluations of programs implemented under real-world conditions and where complex patterns of nesting occur. At least three additional programs participated in the Phase II study according to the minimum criteria above but we were not in sufficient contact with those sites to include them in the evaluation sample. Also, the four Seattle Public Schools programs were very large programs (14 classrooms at four sites), requiring additional assessor visits. For the purposes of the evaluation we collapsed those data into a single rating for each of the four programs. If we counted the number of one-day assessor visits completed for this project, the sample program size is 35 and the total number of ratings is 54.

The sample included systems with QIOs of varied expertise with continuous quality improvement generally and the SLPQI in particular. The QIO in Seattle, School's Out Washington, has extensive experience with continuous quality improvement in afterschool programs and supports a quality improvement system in the region. Programs in Oakland and Grand Rapids had participated in Phase I, however Oakland's QIO, Oakland Unified School District, was unable to complete more than one assessment due to their own time and staffing constraints. The lead organization in Northern California was new to the work.

Procedures

Data collection for the study included the following measures and sources of performance information.

Project records. Project records included records of training attendance, assessor reliability test results, dates for submission of Excel workbooks, dates when performance reports were sent to each of the 32 programs, and notes from technical assistance calls.

Summer Learning PQA Forms A and B. Form A is an observational measure designed to evaluate "point of service" contact between youth and instructional staff. Each observation, morning and afternoon, utilized a "walkthrough" method where the assessors collected systematic anecdotal notes, a detailed running record of staff behavior and youth responses, during 15-30 minute observation blocks in a cross-section of program offerings led by different program staff. Each rating was based on a total of approximately 90-minutes of observation time. Assessors then used the anecdotal records to score the rubrics that constitute Form A, typically requiring about 60-minutes of time to convert the anecdotal records into a complete Form A rating. For full-day programs a distinct Form A rating was produced for the morning and afternoon sessions. For half-day programs only the morning rating was produced.

Form B is an interview-based assessment of management practices. To complete Form B, the assessor interviews the program manager and records written responses. Later this written record is used to score the Form B rubrics, typically requiring about 30 minutes.

Observation of Child Engagement. The Observation of Child Engagement (OCE) is a global measure of child engagement that provides information on the proportion of individuals in a setting that demonstrate five aspects of engagement. The OCE was originally developed for use in early childhood studies (Hamre, Hatfield, Pianta, & Jamil, 2014) and was adapted by the Weikart Center for use in the Texas 21st Century Evaluation in 2011 (Naftzger et al., 2013). The OCE is completed during the Summer Learning PQA Form A observation by rating group behavior at a series of ten-minute intervals during the observation. The maximum number of ratings per observation was 12.

Assessor survey. The assessor survey was developed to better understand successes and challenges in the assessment process and to gain assessor perspective on the Summer Learning PQA. Among the 18 assessors who participated in the Phase II study, 15 responded to the survey (83 percent).

Program manager survey and interviews. The program manager survey was developed to assess a number of attributes at each site, including: (1) fidelity of SLPQI implementation, (2) variations in SLPQI implementation, (3) barriers to implementation, (4) customer satisfaction with SLPQI and Summer Learning PQA and (5) validity of the Summer Learning PQA. Of the 32 participating sites, 14 program managers responded to the online survey (44 percent), representing at least one program from each of the five systems described in Table 2. In the case of Seattle Public Schools programs, the system lead also provided a response.

Phone interviews were also scheduled with an additional nine program managers about SLPQI implementation. Interviews included five questions: (1) Overall, was the SLPQI process worth your time? (2) What were your biggest obstacles with respect to implementing the specific elements of the SLPQI process? (3) How do you plan to use the data report? (4) In the future, what kind of supports would you need for successful implementation of the SLPQI process? (5) Do you have any final thoughts to share about any aspect of the SLPQI process or the Summer Learning PQA tool? All site coordinators we were able to reach consented to speak with us at that time. The total response rate for post-program information about implementation of the SLPQI is 63 percent, or 20 of 32 programs. The issue of missing implementation data is discussed in Appendix D. Table 3 below describes the sample sizes by measure.

Table 3
Sample Sizes by Measure

Measures	Completed Responses
Summer Learning PQA Form A	Completed for all sites (54 total observations; 22 sites with both AM and PM; 10 sites AM only)
Summer Learning PQA Form B	Completed for all sites (32 interviews)
Observation of Child Engagement	Completed for all 32 sites
Site coordinator implementation & satisfaction survey and follow-up interviews with survey non-respondents	Completed for 20 of 32 sites and for 25 unique program managers (15 surveys, 10 interviews)
Assessor implementation and validity survey	Completed for 15 of 18 assessors (15 interviews)

IV. Summary of Findings for Implementation and Customer Satisfaction

In this section we describe evaluative findings, implementation and customer satisfaction. We first describe implementation of SLPQI supports, then fidelity and feasibility of the SLPQI sequence in the 32 programs, and finally, we describe participant satisfaction with the process.

Implementation of SLPQI Supports

SLPQI supports are the training and technical assistance necessary for program managers to implement the work. Since this second phase of the pilot was meant to test the feasibility of implementing the four SLPQI elements, an implementation plan was designed with input from leaders of the five systems noted in Table 1.

Program Manager Training. Site coordinators from both Seattle and Northern California (including the Oakland program) were invited to attend live trainings including the Summer Learning Institute, Instructional Coaching and Planning with Data. The top panel of Table 4 describes training dates and attendance. In general, participants gave the trainings positive ratings (see discussion of customer satisfaction below) and made a number of additional comments. Specifically, participants were positive about the opportunity to meet other program staff and the opportunity to think together about program design. Participant responses included: "the chance to discuss programming with others in the pilot project was important," "[the training was] very relevant and good at forcing me to think about next summer now," and that it was great to "talk about the big picture and general best practices, then compare it to data specifically from our site."

Constructive feedback included concerns related to SLPQI development: The Summer Learning PQA was not fully revised at the time of the Summer Learning Institute trainings so a final version of the standard for quality was not available – several participants commented that this made the process more difficult and our experience would suggest that having the quality measures available to users in advance of an external assessment is critical for building trust in the process. Further, participants felt that the NSLA and Weikart materials were not fully integrated, and in the Summer Learning Institute offered in Seattle, at least one participant reported that the trainers were not fully aligned in their understanding of the project.

Assessor Selection and Training. In Northern California a voluntary application produced a cohort of seven assessors who completed the assessor training. Seattle had eight assessors who were

already reliable on another Program Quality Assessment[9], and they attended a live Summer Learning PQA assessor training. Finally, for the two assessors in Grand Rapids and three in Oakland who had participated in Phase I, the Weikart Center conducted a two-hour webinar to update them on the changes made to both the tool and data collection protocol. The training for all assessors focused on two major content areas: understanding of the measures unique to the Summer Learning PQA and instruction on the data collection protocol. In all cases, assessors were required to achieve 80 percent or greater perfect agreement with gold standard scores of a video-taped program offering.

Assessors reported in training evaluations that the events were worth their time and they acquired new skills or strengthened skills they already had. Participants stated that the trainings provided "Great info," that they liked "Learning new [PQA] components," "Going through the new items… working through the process with others," and that there was "Practice with useful examples."

Constructive feedback included suggestions to have more time to practice with the new scales and items and a suggestion from two participants to find a way to organize a live practice observation. There are opportunities for more video practice and/or live observation pre-work within the current training structure and these issues are typically worked out with the system lead during the design phase at the beginning of the program cycle.

Table 4
Assessor trainings

	Location	Date	Attendance
Program Staff Training Events			
Summer Learning Institute (+ Literacy workshop)	Stockton	April 24-25	41
Summer Learning Institute	Seattle	April 18	24
Quality Instructional Coaching	Stockton	May 30	20
Quality Instructional Coaching	Seattle	June 12	21
Planning with Data	Stockton	August 20	8
Planning with Data	Seattle	August 27	22
Assessor Training Events			
Assessor Reliability Training	Stockton	May 14-15	7
Assessor Reliability Training	Stockton	May 29	7
Assessor Reliability Training	Seattle	May 3	8
Assessor Reliability Webinar	Grand Rap/Oakland	June 11	4

---

[9] The Summer Learning Program Quality Assessment is similar to the Youth Program Quality Assessment (Youth PQA) and School Age Program Quality Assessments (School Age PQA) in that it shares many of the same items and uses a similar method for collecting evidence and scoring the instruments rubrics.

**SLPQI Implementation Fidelity and Feasibility**

To learn about implementation of the SLPQI, program managers were asked about relative ease or difficulty they experienced coordinating the four SLPQI elements, specific issues or obstacles that arose during the sequence and the overall success of the improvement cycle. Assessors were asked questions about their ability to implement the specific aspects of the intervention in a timely fashion.

Implementation Fidelity. To assess the level of implementation at the 14 programs with complete program manager survey responses, an implementation index was created to describe overall implementation of the four SLPQI elements. The index presented in Figure 2 was created from responses to four items (1=implemented in program, 0=not implemented) and ranges between 0 and 4. The items were: Program staff participated in intervention supports (e.g., training); the Summer Learning PQA assessment was completed; the program manager provided instructional coaching to staff; and the program staff engaged in improvement planning with their SLPQI performance report.

Over half of the responding program managers were able to complete three or more of the four implementation elements included in the index. Ten of 14 respondents stated they attended the Summer Learning Institute. Nine stated they were able to coach their staff on quality instructional practices using the Summer Learning PQA as a guide. Five of 14 respondents stated they were able to work with a team of staff to create an improvement plan based on the results of the Summer Learning PQA.

While these results are of interest, they do not of themselves explain whether or not implementation fidelity for the SLPQI was good or bad. To offer an answer to this question, we explore the context of implementation and then examine implementation in the two systems with the most supportive context – where we would expect the highest levels of implementation to occur.

Figure 2
SLPQI Implementation Index
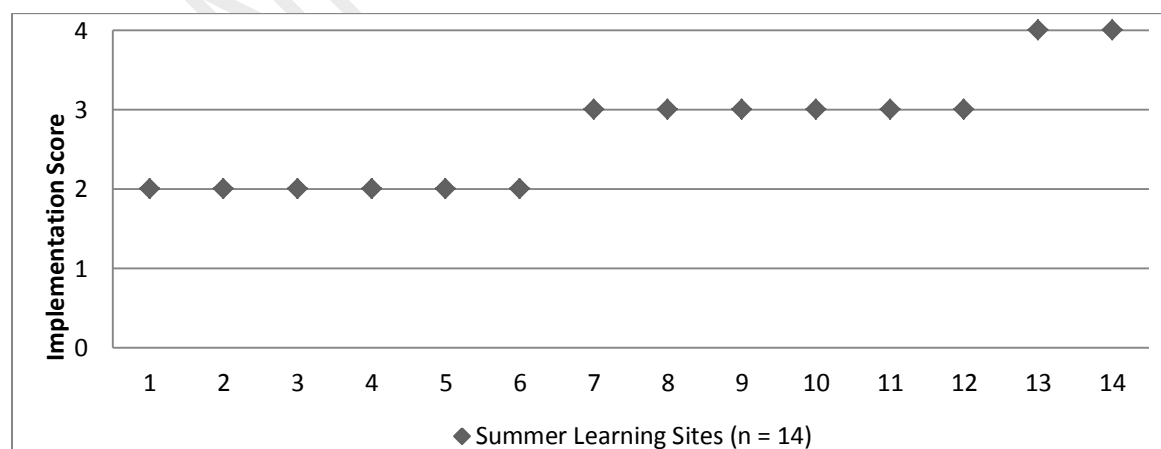
Context of implementation. What we know from these and other sources of project data is that (1) most study sites were represented at the Summer Learning Institute and Instructional Coaching trainings; (2) that all study programs received an assessment and were sent a performance report and (3) that project correspondence occurred during the project with all 32 program managers. However, in the Northern California cohort there was no existing quality improvement infrastructure and few prior connections between the programs. The QIO, the University of the Pacific in Stockton, CA, was able to support recruitment and logistics and but did not have prior relationships with the cohort of programs. This lack of either prior collaborative experience or formal relationships made all aspects of implementation more tenuous and required more direct communication between the NSLA/Weikart project team, the programs and the assessors.

In particular, scheduling assessments in a timely fashion was a challenge in the Northern California sites. Table 5 provides a description of when SLPQI data was received by the Weikart Center from assessors and when the performance report was sent back out to the program managers. In most cases data was received and the performance report returned in a less than two day turnaround, often next day. However, Table 5 also indicates that the assessor data was received at the Weikart Center after the program's end date in five of the 11 cases and in two additional cases the performance report was sent back to the program manager within two weeks of the program's end date. None of the five Northern California sites with a post-project survey or interview response reported completing an improvement plan with their staff team during the program session (although one reported that improvement planning would occur for next year). Table 5 describes part of the reason why it was difficult to engage the staff team in improvement planning during the summer session in the Northern California programs. In the majority of cases the staff team was either gone or wrapping up the program by the time the performance report was received.

In contrast, the two Seattle systems (Seattle Public Schools and City of Seattle) had a more supportive implementation context. The two Seattle cohorts included a substantial number of programs that had been using the standard YPQI for several years, a pool of experienced PQA assessors and perhaps most importantly, a high capacity QIO (School's Out Washington) supporting the project. Most Seattle based programs simply extended their existing school year quality improvement capacity to accommodate the SLPQI sequence during the summer months.

Table 6 describes SLPQI fidelity in the 11 Seattle programs that returned implementation surveys. In general, implementation of the SLPQI was mixed across the Seattle sites with all sites receiving the assessment visit, most attending the Summer Learning Institute and coaching their staff, and only four of 11 producing an improvement plan with a program team during the summer session. While we would describe this level of implementation as mixed, or a moderate level of fidelity, we provide

Table 7 as a benchmark for comparison. Table 7 presents a summary of implementation of the four

SLPQI practices alongside similar figures for the Youth Program Quality Intervention study conducted in

89 afterschool programs in four states in 2008 (Smith et al., 20012, see Appendix Table E-3). Note that

implementation fidelity for the YPQI study treatment group is comparable to the SLPQI group while

YPQI Study control sites implemented at a much lower level than the SLPQI group.

Table 5
Northern California Program Dates for Performance Report Turnaround and Program End Date

	Program End Date	Data Received	Data Packet Sent to Program
Child Development Centers – Jacobson 21	8/1/14	6/22/2014	6/24/14
City of Richmond – Kids N Motion: Nevin	8/9/14	7/12/2014	7/14/14
City of Richmond – Kids N Motion: Shields-Reid	7/11/14	7/15/2014	7/16/2014
Housing Authority of San Joaquin - Conway	7/22/14	7/13/2014	7/14/14
Housing Authority of San Joaquin – Sierra Vista	7/23/14	7/8/2014	7/8/14
Stockton - Harrison	6/20/2014	7/2/2014	7/2/2014
Stockton - Pittman	6/20/2014	7/1/2014	7/1/2014
YMCA of San Joaquin – Lincoln HS	8/22/2014	7/24/2014	7/31/2014
YMCA of San Joaquin – McKinley Park Day Camp	7/25/2014	8/1/2014	8/4/2014
Stockton Unified – Roosevelt School	8/2/2014	7/3/2014	7/14/2014
Stockton Unified – Freemont School	6/20/2014	7/1/2014	7/2/2014

Table 6
Implementation Index Data - Seattle

	Summer Institute	Assessor Visit	Coach Staff	Improvement Plan
Northgate ES	1	1	0	0
Eckstein MS	1	1	1	0
Mercer MS	1	1	0	0
Aki Kurose	1	1	1	0
Chief Stealth	1	1	1	0
America SCORES	1	1	1	1
Madrona K-8	0	1	1	1
Cleveland HS	0	1	1	0
MLK Summer Staircase	0	1	1	0
Franklin HS	1	1	1	1
Seattle World School	1	1	Missing	1

Table 7

Comparison of Seattle Implementation Index with YPQI Study Treatment and Control Groups After One Intervention Cycle

	Implementation Index for 11 Seattle Sites	YPQI Study Treatment	Control
% sites 1 practice	0	4	40
% sites 2 practices	36	13	34
% sites 3 practices	54	32	10
% sites 4 practices	9	53	16

Challenges for Implementation. When asked about obstacles to successful implementation, all site coordinators we interviewed cited timing as the primary challenge to overcome, especially in programs that employed school-day teachers. For example, the Summer Learning Institute was a challenge to schedule in the short window between the end of the school-year and the beginning of the summer session. Scheduling the assessments was also difficult and several program managers noted that assessments need to be strategically scheduled: too early, and the program has not completed the intake process with their students, as one program noted; and if the assessment was scheduled too late, the performance report might come too late to be used with program staff. Some assessors were unable to schedule a visit until almost the end of the program session. In Seattle, the Planning with Data workshop occurred the same week that programs were wrapping up and shifting into the school year. The network lead reported that it was challenging for participants to fully engage with the improvement planning process as they were physically and mentally exhausted from running the summer program and their thoughts were moving to the school year.

While time was a challenge, eight of 14 program managers also stated that the trainings, including the Summer Learning Institute, were well-timed within the framework of the SLPQI. It appears that while timing was not a problem for all programs or staff, timing difficulties were consistently experienced with specific actors (e.g., school teachers in Seattle) and specific aspects of the SLPQI (e.g., scheduling assessments in Northern California). Indeed, the desire for additional training was expressed by five of the nine program managers interviewed, specifically in reference to training for school-day teachers and new staff.

Another important set of challenges noted by both program managers and assessors occurred around the role of the assessor and the performance report. One challenge was again related to timing: first-time assessors in particular struggled with submitting their data within the 48-hour time period, which meant that sites were not able to receive their data in a timely fashion. In several cases, late data submission by the assessor led to late delivery of the performance report to the site coordinator.

A second challenge around the assessors and the performance reports was related to our decision to automate the performance reports in order to improve the consistency of the report quality. In at least one case the assessor thought the performance report was invalid and sometimes contradicted the program's strengths. Assessors articulated a desire to have more control over producing the content of summary report.

A question to explore with future iterations is whether the assessors should not only have more autonomy in the interpretation of the performance report but also should have an expanded role in presenting the performance report to the program manager as an assessor-coach. Having the assessor extend their role to include coaching around interpretation of the performance report may help to build an effective learning community around the data, and relieve some burden of the program manager to communicate the findings to staff.

Several additional implementation challenges were mentioned by one or more program managers and/or assessors: First, system leadership and especially access to a high capacity QIO is an important support for successful implementation. Although the University of the Pacific in Stockton, CA played a crucial role in organizing the logistics of each training and helping to communicate the schedule, they did not have direct ties to all of the participating sites which created a disjointed system. This lack of cohesion may have contributed to the challenges scheduling assessments and the low attendance at the Planning with Data training in August 2014.

Second, while the Planning with Data workshop is designed to support improvement planning, system leads and program managers reported needing more explicit structure and concrete suggestions for how to conduct improvement planning during the summer session and how to design improvement plans that could stretch from one summer to the next.

Third, programs that were experienced with the standard YPQI felt some dissonance with the content offered by NSLA staff during the Summer Learning Institute. It is likely that this disconnect was partially due to the fact that the Summer Learning PQA had not been finished at the time of the training, but it is also indicative of a need for Weikart and NSLA to make explicit connections and alignment with quality improvement content.

Feasibility. To address the question of feasibility we asked site coordinators if they felt they were able to successfully implement the SLPQI at their sites. Twenty-three of 24 program managers who responded to this question during a survey or interview stated that they were able to implement the SLPQI successfully. One Seattle respondent felt some of their success was due to good existing relationships between the QIO and participating programs. Two respondents credited "good communication" among program participants, one noted the site's practice of a "daily debrief" where participants discussed the implementation as it was happening and ways they could support the quality improvement process. Two

respondents credited their success with previous exposure to the YPQI and one stated that the SLPQI process "felt like a natural transition for us."

Assessors were also asked to describe the ease with which they were able to carry out the related assessment elements. Table 8 presents the assessor feedback on feasibility of several primary tasks of the assessor. We also present responses for experienced assessors to examine our assumption that SLPQI methods get easier to implement with practice. Contrary to our expectations, experienced assessors who had participated in the SLPQI during a prior year tended to rate task feasibility as more difficult.

Table 8
Feasibility of Assessor Tasks

Assessor Tasks	All Assessors (1= hard, 3=easy) (n=15)	Experienced Assessors Only (n=7)
Conduct the observation	1.70	1.71
Score Form A	1.40	1.28
Conduct the interview	2.50	2.42
Score Form B	2.00	1.85
Use the Excel workbook	2.00	1.85

Customer Satisfaction

Satisfaction with the Summer Learning PQA. Both program managers and assessors were asked about the adequacy of the Form A academic items as representing high quality academic instruction and the ability of the overall quality score to differentiate between programs that were of differing quality.

In general, the responses were positive about the standard for quality of academic instruction: 11 of 14 program managers who responded to this item stated that the Summer Learning PQA accurately identified the quality of academic practices at their sites – with the remaining three citing implementation failures as the reason for their dissatisfaction with the new academic rubrics (e.g., didn't receive performance report at all, didn't receive report before end of program, assessor visited on a bad day). Assessors were also asked if they felt the Summer Learning PQA tool was able to accurately assess the presence of academic practices. All 15 assessors stated that the tool was an accurate assessment of academic practices.

One site coordinator stated that they were "very impressed with the math and literacy items because they address social-emotional and not just specific math or reading proficiency skills" and that "These items should be added to the Youth PQA." Another stated that the Summer Learning PQA "framework captures the unique challenges of integrating interaction/engagement while maintaining a high level of academic instruction."

As an assessment of overall quality, 10 of 14 stated that the Summer Learning PQA provided an accurate assessment of overall quality practices. All 15 assessor respondents stated that the Summer Learning PQA scores differentiated between programs of higher and lower quality. One site coordinator among the Northern California sites felt strongly enough about the measure to ask if it could be included in his curriculum for an assessment class he was teaching at the local community college.

Overall Satisfaction with the SLPQI. With respect to overall satisfaction with the SLPQI, 21 of 23 program managers agreed with the statement that it was "a good use of their time and effort" to implement the SLPQI. One site coordinator stated that the SLPQI process was both "worthwhile and helpful" for providing programming guidance: "It was worthwhile. It helped us develop a stronger focus for this summer." Another site coordinator stated, "The assessment helped us find areas that we could strengthen for next year." Another site coordinator stated, "As a team we took a collaborative approach and were very intentional about doing this. I thought it turned out well." One site coordinator stressed the value of the quality data provided by the Summer Learning PQA: "the data was very valuable in that it gave us a way to look at the program we delivered against other program standards." And, "It present[ed] interesting information we will consider as we think about the summer program again." One site coordinator identified the trainings as particularly helpful: "I received a deeper understanding of 'Quality Coaching.' It's vital as supervisors and staff to understand the deeper purpose of 'Respect' 'Observe' and 'Support.'"

The one respondent indicating that it was not a good use of time stated that he had not received his data report in time to make course corrections. Another site coordinator demonstrated how time challenges for use of the performance report during the summer session could coincide with judgment that it was still worth the effort to support planning for next summer's cycle: "…we're only a four week program, it would have been hard to do a full stop and implement changes anyway… I plan on planning [for next summer] regardless. What I got from the report, if we'd had consistent staff, I would love to use it as a training tool but since we will have none of the people who were with me last summer, I will be the only repeat factor so it will most heavily weigh on the planning."

Table 9 provides customer satisfaction data collected during eight SLPQI training events where ratings range between 1 and 5 with 1 indicating strong disagreement with the statement and 5 indicating strong agreement. All training types received generally high satisfaction levels with the exception of the workshop during which the improvement plan is developed, where participants generally approved of the training content but some doubted the likelihood of implementation or adequacy of administrative support.

Table 9

Customer Satisfaction Data from Eight Training Events

	Summer Learning Institutes on 4/18, 4/24-25 (N=18)	Assessor Trainings on 5/3, 5/14-15 , 5/29 (N=22)	Instructional Coaching on 5/30, 6/12 (N=30)	Planning with Data on 8/20, 8/27 (N=13)
Information provided was of high quality	4.67	4.56	4.6	4.31
Workshop materials were useful.	4.57	4.36	4.53	4.31
I feel that today's workshop was worth my time.	4.57	4.56	4.33	3.69
I feel that within the next 30 days, I will implement the information or skills that I learned.	4.61	4.53	4.53	3.62
I feel that the content is applicable to my current job or position.	4.47	4.80	4.47	4.38
I feel that I have administrative support at my program to implement the content.	4.42	4.14	4.30	3.54

V. Results for Quality of Summer Programs

Implementation of SLPQI in the 32 study programs also produced detailed performance information about the quality of management and instructional practice. Because few summer learning programs have been examined at this level of granularity – measurement of specific managerial and instructional practices – this section provides an aggregate description of summer learning practices using SLPQI data. These findings reflect the type of secondary data product that SLPQI systems can produce as an aggregate description of quality across many individual programs. The Summer Learning PQA Form A includes eight domains composed from18 scales and descriptive data at the item, scale and domain level are provided in Appendix A.

Quality of Instructional Practices

This section presents performance information based on Form A data, in some cases differentiating between an *active-participatory* definition of quality practice and an *academic* definition of quality practice, and in some cases presenting a composite Instructional Total Score that combines measures across domains. The Pearson-r correlation coefficient for the AM and PM sessions for the Instructional Total Score was r=0.59[10] for the 22 programs with both AM and PM ratings.

Active-Participatory Instruction. Figure 3 shows average quality ratings for four Form A domains that represent a definition of quality derived from an *active-participatory* instructional method that structures the standard Youth PQA (Smith & Hohmann, 2005). Overall, the 32 summer learning programs demonstrated a profile of practice with a very similar shape to profiles generated from the Weikart Center's normative data bases.

Academic Instruction. Four additional Form A scales were constructed as part of Phase II study - Learning Strategies, Higher Order Thinking, Math, and Literacy - which reflect an academically focused instructional method. Figure 4 presents the profile of quality scores for these four scales, again comparing the mean scores for morning and afternoon sessions. There was a consistent pattern of differences between morning and afternoon programming on the *academic quality* scores. For each of those scales, mean scores were substantially higher in the morning (See Figure 5). Figure 5 presents the Instructional Total Score, an average across all Form A Domains (except Safety, Math, and Literacy), by morning and afternoon session for each program, again indicating that interpretation about program level performance can only be made from program level data (i.e., system averages cannot be used to describe individual

---

[10] None of the correlations presented in this report have been disattenuated to correct for measurement error that is certainly present, suggesting that these correlation coefficients represent the lower bound for a true score correlation that would be higher.

sites). As a rough guideline, programs with Instructional Total Scores below 2.9 can be considered low quality[11], as are seven of the 22 programs in the study with both morning and afternoon data. Program with an Instructional Total Score above 4.1 can be considered high quality and this would include the seven highest scoring programs in Figure 7. Note that high and low quality defined in this way denote roughly the bottom and top quartiles in the sample.

Figure 3
Average Scores for Active-Participatory Quality in AM and PM Sessions

---
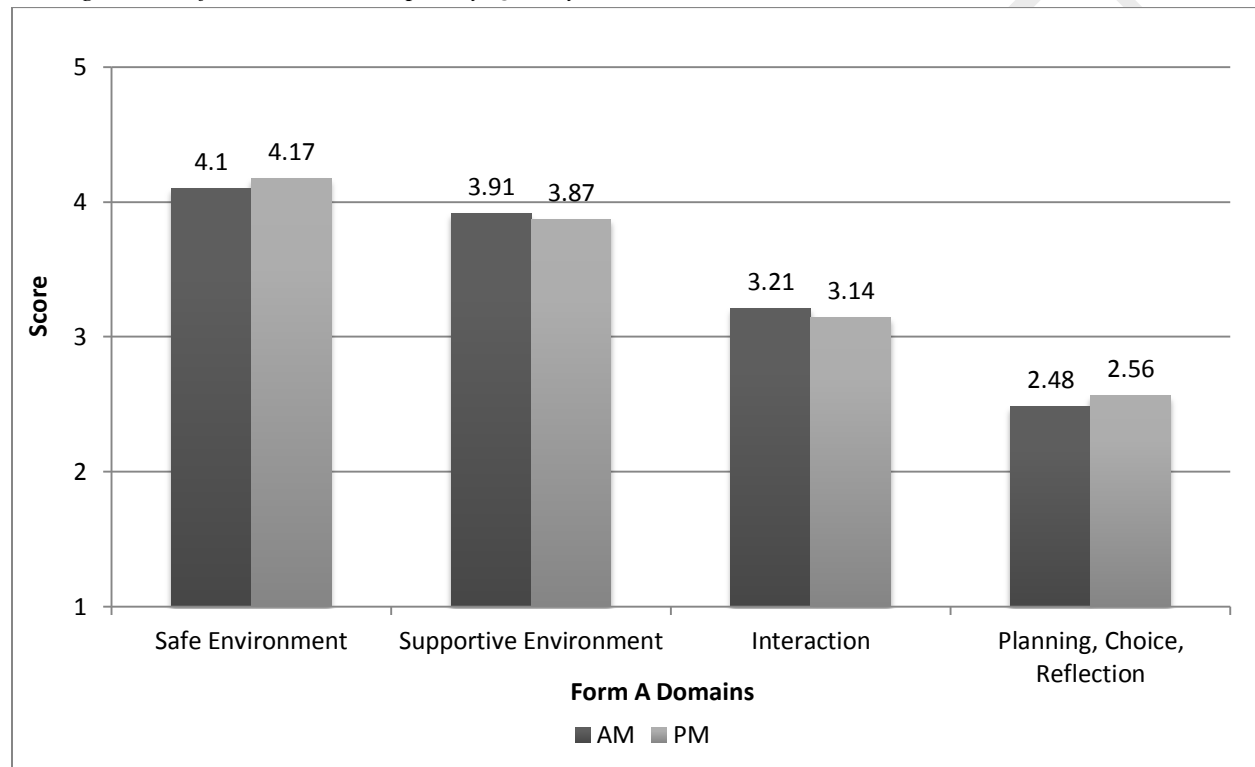
[11] Based on a recent national convenience sample of Youth PQA data, collected by trained external assessors (n= 505), Instructional Total Scores below 2.87 were in the bottom quartile of all scores. Previously established thresholds of quality for the Youth Program Quality Assessment (Youth PQA) identified instructional total scores (the average total score of the Support, Interaction and Engagement domains) below 2.9 as indicative of low levels of interest, challenge and belonging for the young people in those settings (Akiva et al., 2010) . More recent norming work based on multiple years of PQA data in Palm Beach County, Fla. designate an Instructional Total Score of 4.1 as the bottom of the high quality category (Smith, 2013).

Figure 4
*A*verage Scores for Academic Quality in, AM and PM Sessions
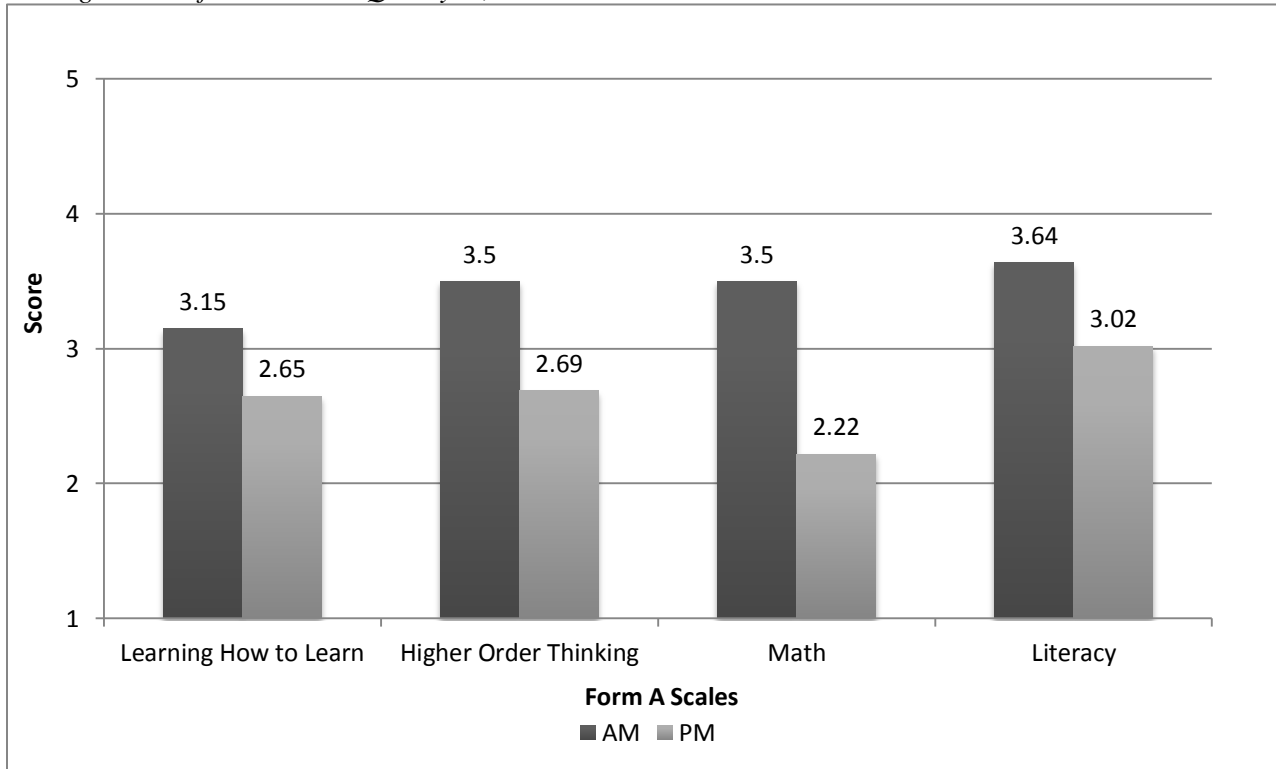
Figure 5
Instructional Total Scores by Program for AM and PM
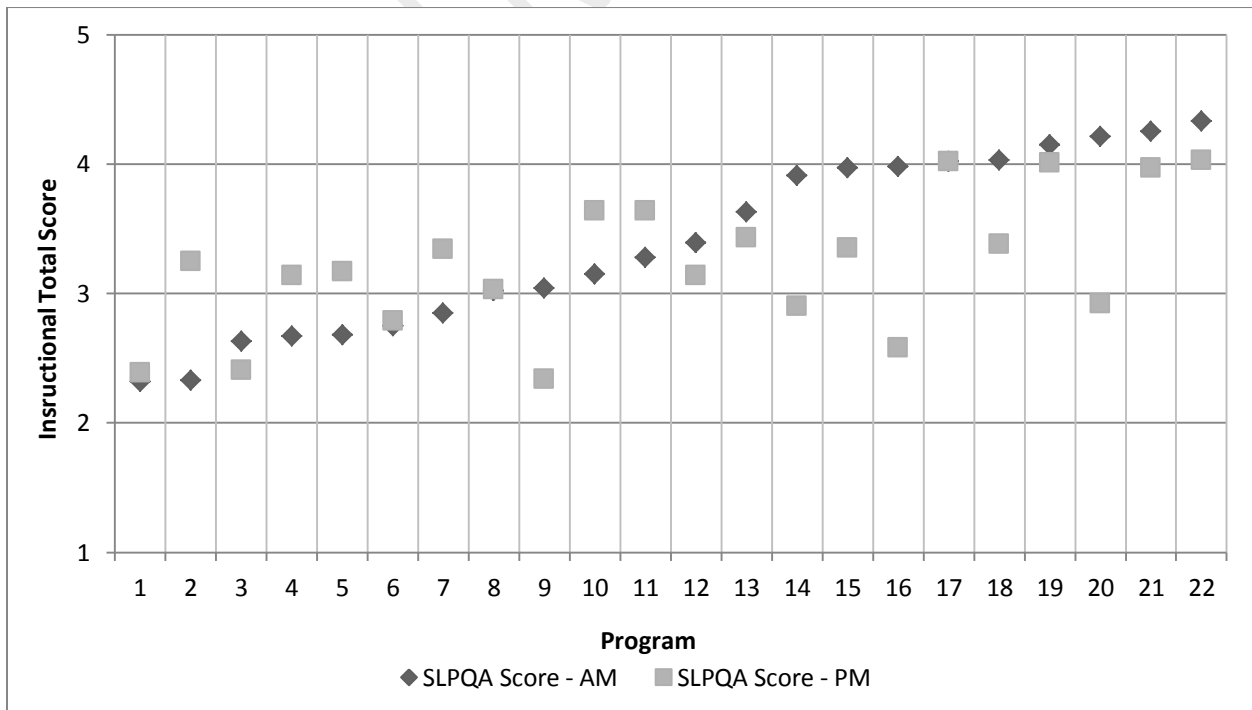
Finally, Table 10 presents a selection of lowest scoring items across all eight Form A domains. Across 44 distinct morning or afternoon quality ratings, these staff practices which were not present during 40 percent or more of those ratings, i.e., the practices are rare.

Table 10
Low Scoring Items: Summer Learning PQA Form A , N=44 Ratings

Item	% scoring 1
Opportunities to make plans	52.3
Content alternatives	45.5
Process alternatives	40.9
Intentional reflection	52.3
Structured opportunities to provide feedback	59.1
Staff encourages youth to deepen knowledge	45.5

Quality of Management Practices

The Summer Learning PQA Form B includes four domains including planning, staff training, family connection and individualization. Figure 6 provides domain averages for all 32 sites in the study sample. Figure 7 provides the Form B total score for management practices, a mean score across the four domains, by the Form A Instructional Total Score to present a profile of site quality in terms of management practices and instructional practices. For the sample of 32 sites these two composite scores have a Pearson-r correlation coefficient of r=0.5.

Figure 6
Average Scores for Management Quality
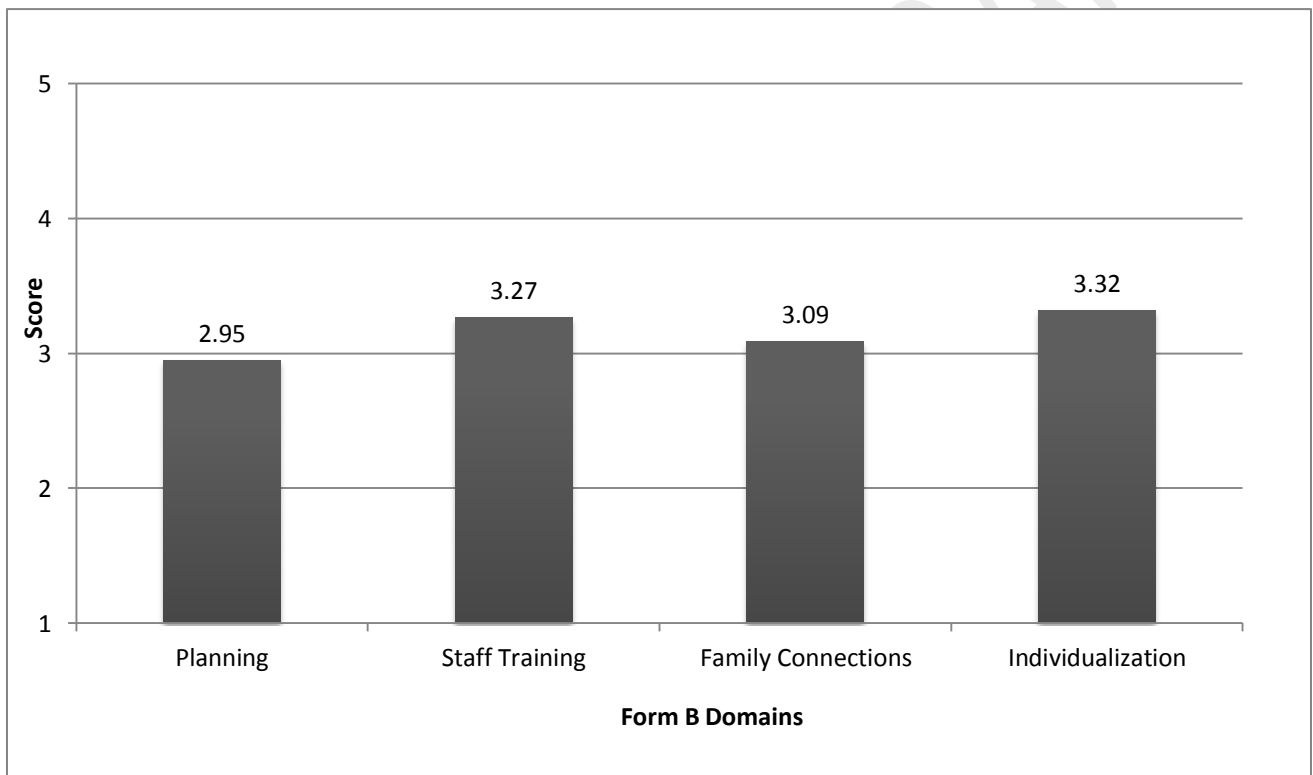
Figure 7
Management and Instructional Quality by Program
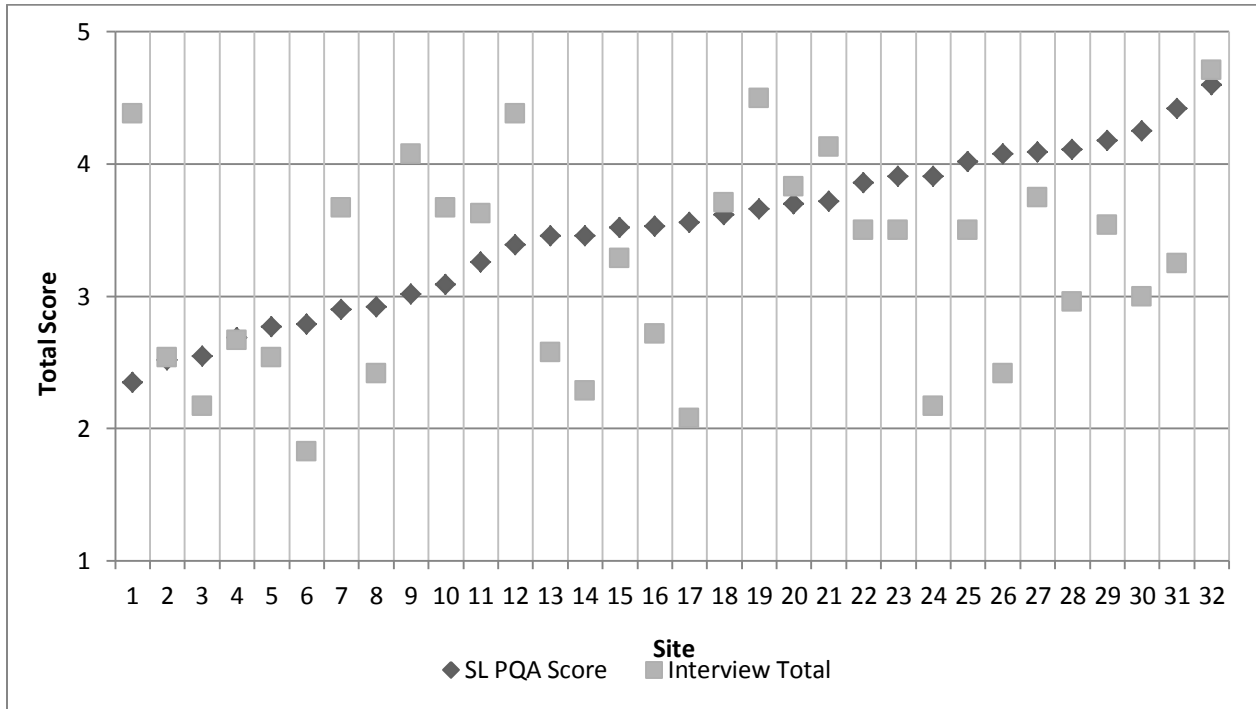
Child Engagement

Assessors in the Phase II study also completed a global assessment of child engagement every 10 minutes during the morning and afternoon observation periods. Figures 8 and 9 suggest that child engagement was high in the morning and mid-afternoon, with a dip before and after lunch and a more pronounced dip towards the end of the program day. Overall levels of engagement were lower in the afternoon.

Figure 8
Average Percentage of Students Identified as Mostly or All Engaged – AM Session
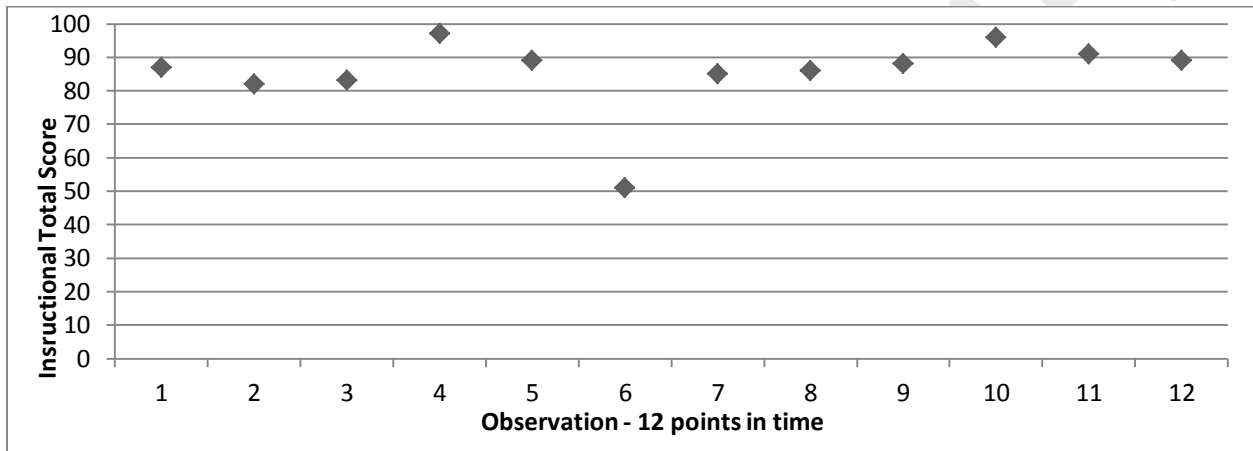
Figure 9
Average Percentage of Students Identified as Mostly or All Engaged – PM Session.
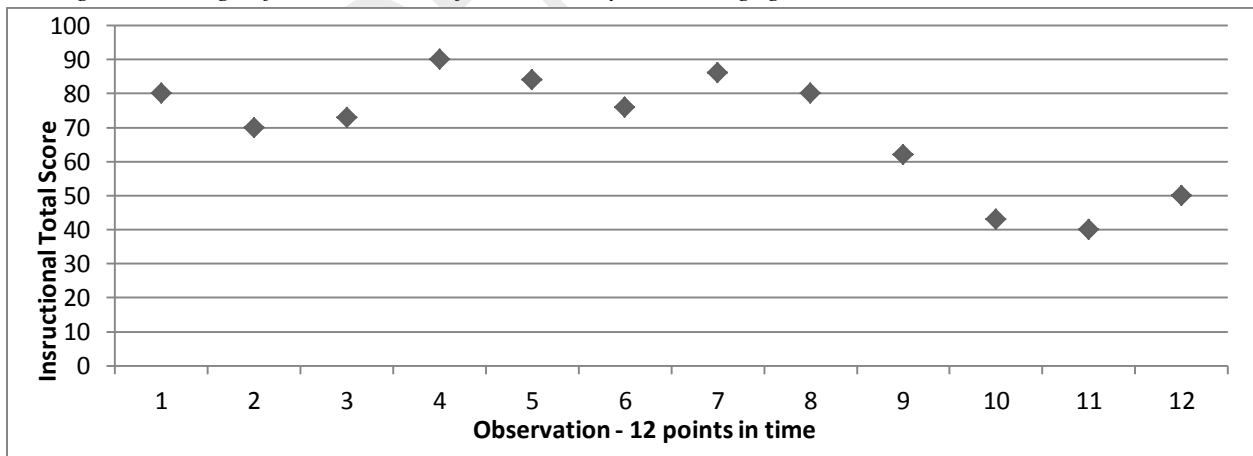
Greetings, Transitions and Departures

Assessors also completed a checklist related to basic best practices for three transition periods in the program day. Ratings for the Greetings Index were only collected during morning observations while ratings for the Departures Index were collected only during afternoon observations. Ratings for Activity Transitions Index were collected during all observation periods. Table 11 suggests that, like school day classrooms, transitions are a challenging part of summer programming. Somewhat alarmingly, these data indicate that basic best practices related to children's sense of belonging (welcoming, culture) are not deployed consistently in a majority of programs while in 25 percent of afternoon offerings children are left unattended for at least some period of time.

Table 11

Percentage of Programs Implementing Specific Transition Practices

Index Item	% of Ratings, N=44
Greetings: Opening and arrival time	
Children greeted by staff	71.0
Session starts within 10 minutes of scheduled time	72.7
Welcoming activity or icebreaker	39.4
Incorporates themes or aspects of program culture	45.5
Activity Transitions: When a group of children moves to another activity in a different location	
Smooth and quick transition times	85.4
Clear transition communication	91.6
On task and ready for transition	83.1
Activity choices clearly communicated	84.4
Program lessons incorporated	22.6
Departure: When children leave for the day.	
Organized process	71.9
Smooth process	78.1
Constructive activities while waiting	46.9
Children left unattended	25.0
Utilizes parent engagement opportunity	43.8
Verification system	50.1
Program incorporated	9.7

Note: Percentages refer to the percentage of all observations where this quality practice was observed.

VI. Discussion and Recommendations

Effective continuous improvement, accountability or quality assurance systems can also be effective adult learning systems. In education settings, the continuous improvement work requires a short list of active ingredients – a professional learning community connecting performance feedback to improvement action while building capacity and expertise over multiple cycles. The SLPQI was designed to support effective professional learning communities with four elements: (1) a shared standard/measure for management and instructional practice, (2) training and technical assistance supports for implementation of the continuous improvement elements, (3) data products that provide timely performance feedback on objective behaviors aligned with specific roles and (4) empowerment of teams to set their own quality improvement agenda based on an improvement cycle that is flexible enough for local adaptation but sufficiently structured to promote timely receipt of supports.

Our experience suggests that the system-level design for implementation of the continuous improvement elements will be different in each case at each level. Given this design principal, our task is to (1) support system-level and organization-level actors to innovate on the path to implementation of the core elements and (2) to continuously improve and evaluate the core elements in various design configurations.

This study represents this type of evaluative effort and sought to answer the following questions: How feasible was implementation SLPQI? How can the feasibility be improved? Was the SLPQI useful and therefore a good use of time? Did the Summer Learning PQA identify areas of high and low quality?

Findings

The following findings were produced through the Phase II Feasibility Study:

1) ***Overall satisfaction with SLPQI and Summer Learning PQA was high***. Nearly all respondents gave a positive endorsement to their participation in the SLPQI, indicating the SLPQI was a good use of their time and that it was feasible to implement. Further, most site coordinators and all assessors said that the Summer Learning PQA was useful as a standard, and scores described real differences in the quality of programs.

2) ***Implementation of SLPQI was uneven across sites, although partially by design***. Few study participants achieved full implementation of all SLPQI elements. Nearly all program managers participated in pre-program trainings and all sites received an assessment of management and instructional practices by a trained assessor. Despite challenges with missing data on some aspects of implementation data, we estimate that over two-thirds of site coordinators reported coaching staff on instructional practices and over one-third of sites created a program improvement plan based on SLPQI performance data. Overall rates of implementation are

comparable to those seen in other studies of similar interventions and may also reflect a strength of the continuous improvement model: Innovation in the implementation of core elements is required to achieve optimal fit and programs typically learn to implement in successive program cycles.

3) ***Readiness for SLPQI can be characterized by several key features and time is the greatest challenge***. Readiness to implement SLPQI is increased when: (a) the system leads engage with an experienced technical assistance provider to clarify purposes and resources necessary to support an explicit design for the organization-level continuous improvement cycle and system-level supports; (b) an experienced quality intermediary organization (QIO) provides timely supports and coordination according to the design; and (c) experienced assessors are available for the full cycle of data collection and reporting. There was higher implementation fidelity in places where roles and responsibilities were clearly defined and assigned. Timing was the most significant challenge to implementation for training attendance, scheduling assessment visits, and conducting team-based improvement planning.

4) ***Precision and meaningfulness of Summer Learning PQA data is promising***. The Summer Learning PQA Form A was endorsed by program managers and assessors as effectively describing high quality instructional practices and differentiating between programs of high and low quality. Several reliability analyses suggest that the Form A Instructional Total Score demonstrates adequate consistency across raters and across short time periods; and when multiple ratings in the same site are combined as a composite score sufficient consistency within organizations to produce a program-level quality rating. Validity evidence suggests that the Form A scores are associated in the expected direction with several important characteristics of summer learning programs.

5) ***The sample of 32 summer learning programs contained a mix of higher and lower quality features, with differing profiles for active-participatory versus academic instructional quality.*** Summer learning programs in the study sample demonstrated wide variation in quality and high quality management practices and high quality instructional practices tended to occur in the same programs. As expected, academically focused sessions tended to score higher than enrichment sessions on measures of academic instructional quality, and child engagement tended to be higher in academic sessions. Approximately one-quarter of programs could be rated as very low quality and one-quarter could be rated as high quality – although more analyses could be done to refine these conclusions. The quality of transitions into, during and out of the program could be substantially improved.

Recommendations

The sequence of design, iteration, and validation toward development of effective interventions and/or precise measures is long and expensive. The Phase II study has advanced our understanding of how continuous improvement elements fit into summer learning programs and has provided an opportunity to continuously improve the SLPQI and supporting infrastructure to a point where implementation at greater scale is possible. While the design, iteration and validation sequence is far from complete, we do believe that we have advanced toward a point where market forces will start to fund that sequence.

As a result of the Phase II study, we offer the following recommendations for next steps in the effort to bring a continuous improvement culture to the summer learning field.[12]

Continue to improve the standard/measures. We have secured funding from the Packard Foundation to further develop the Form B interview by adding new items, cleaning up rubrics, formatting with a layout that matches Form A and adapting the on-line Scores Reporter to include the Form B data in automated reports.

Although development of a more comprehensive Form B will be an important part of improving the Summer Learning PQA measure, the pressing need at this point is to mount a study that further addresses the critical issues of precision (reliability) and meaningfulness (validity) of the data produced by both Forms A and B. In particular, this is the case for the Summer Learning PQA Form A which, as a measure of instructional practice, is in the highest demand and of greatest potential consequence for the field. We need to know if higher quality programs build academic skills during the program period – and at entry to the next academic year. We also need to know if higher quality programs build social and emotional skills that support school year success. The good news in this regard is that many summer learning systems already collect pre- and post-test data on specific academic skills that could be associated with quality ratings. There is interest from school districts in participating in this type of study. For example, we have had preliminary discussions with Seattle Public Schools about siting a study in their 30 summer programs and have heard interest from several other systems as well.

---

[12] Our staff team has come away from this work with the sense that, despite the considerable professional skills of summer program staff, the summer learning programs themselves are often fragmented and impersonal, with insufficient resources dedicated to either sufficient program planning, or to quality of staff relationships and shared purpose. In our interviews, site coordinators seemed harried and staffing inconsistent. Unfortunately, it is also our experience that when staff skills and relationships are undervalued, the same is true for child and youth experience in the setting. A continuous improvement culture parallels the program culture that we want for children and youth.

Further validation work is a worthwhile investment because the Form A measure also serves as a set of standards for good practice in summer learning programs. The design of the Form A rubrics is specifically intended to produce communication value that is critical both for new staff who are teaching in the summer for the first time and for more experienced staff who are trying to more artfully craft child and youth experiences that blend academic content and active-participatory learning. If it is true that what gets measured matters, then we want to make sure that we are measuring the right things. Finally, because the Form A is an observational measure of instruction, it is likely to fit effectively in school districts responding to the growing emphasis on teacher evaluation and observational measurement.

Improve intervention supports. A second primary recommendation is to complete development of support materials for the SLPQI. High quality supports increase the ability of local leaders to innovate unique configurations of the continuous improvement elements in ways that fit local circumstances.

An effective manual is an important support for any intervention and during the Phase II study a number of important pieces of content were developed to support both implementation of SLPQI by site coordinators and data collection and reporting by the assessors. Again, the Packard Foundation has funded development of an SLPQI guide which will explain both the intervention and the methods for using the Summer Learning PQA measures.

Another area for improvement is the training supports for SLPQI users. NSLA and Weikart Center will need to further refine the series of supports that comprise the intervention process, namely the Summer Learning Institute, Quality Instructional Coaching, Summer Learning PQA Assessor and Planning with Data trainings. In particular, the structure of the Summer Learning Institute and Planning with Data need to directly align not only with the Summer Learning PQA measures but with one another as well. They are the anchors at either end of the user experience with the process.

Meeting the summer program stakeholders (i.e. system leads, assessors, program managers and staff) where they are in terms of familiarity with assessment tools and processes is always going to be fundamental in shaping a scope of work with individual client systems. Part of any system-level technical assistance will be mapping the available professional development resources to the needs and resources of each client. Working with the network to determine the best possible timing of training supports will also be an ongoing, evolving process. Weikart Center and NSLA expect to get better at prescriptive recommendations by working through the process with more clients as innovation partners.

Develop the role of the assessor-coach. Clear parameters and expectations around the assessor role are critical to successful implementation of the Summer Learning PQA and production of the performance report. Expanding the assessor's role to encompass one of a technical assistance coach could serve to increase the assessor's stake in the process and sense of accountability to the program.

In the future, we will help summer learning programs develop an assessor-coach role where assessors also provide feedback and potentially play a role in improvement planning. Specifically, the assessor-coach role differs from the instructional coaching content that is currently a part of SLPQI and which is directed at building the abilities of program managers to coach staff during and following instruction. In this case, the coaching method is more like technical assistance in that the assessor-coach needs to be able to join with the program manager to strategize with the about how the SLPQI sequence can be deployed and to interpret the subsequent performance report. In many systems, the assessor-coach may be a primary source of continuity from year to year as summer program staff change – and a source of partnership for summer learning program leads who often work with brand new staff and brand new programs each year.

Improve data products. One key source of improvement for the SLPQI performance report will be the availability of a Summer Learning PQA report through the Weikart Center's Online Scores Reporter. This technology will be fully available for Summer Learning PQA data collection and reporting in the summer of 2016, and Packard Foundation has committed a small sum to development of a basic form for Summer Learning PQA Forms A and B and requisite reporting levels.

Another potential improvement to the SLPQI performance report would be inclusion of child data so that the report would include both setting quality and individual youth data describing the skill baseline and/or growth. This kind of performance report would serve both evaluative as well as positioning purposes in communities and systems. While a potentially useful supplement to the SLPQI data package, child-level data can be specific to networks/districts/states and as such its inclusion must be considered on a case by case basis.

Develop a multi-tiered TA model. Finally, Weikart and NSLA will develop a multi-tiered implementation model to vary intensity and cost across the full range of summer learning system and program needs, from a single program to a regional network. Because it takes multiple cycles to achieve full implementation, it is important to have intervention supports that fit client strategies to conduct pilots and/or build capacity over time. This multi-tiered TA model is described under separate cover as part of Weikart Center and NSLA's organizational planning.

Finally, we believe that school districts may be interested in an intensive training for the assessor-coach role – a district or local intermediary staff person who would become an expert at both Summer Learning PQA assessment and at coaching sites in the use of their assessment data. School districts with mandatory summer school might be interested in this role as an ongoing source of continuity for summer programs where staff and students change frequently.

References

Akiva, T. (2007). *Quality coaching.* Ypsilanti, MI: High/Scope Press.

Akiva, T. (2012). *The psychology of youth participation in organized activities.* University of Illinois.

Akiva, T., Cortina, K.S., Eccles, J.S., & Smith, C. (2013). Youth belonging and cognitive engagement in organized activities: A large-scale field study. *Journal of Applied Developmental Psychology, 34*(5), 208-218. doi: http://dx.doi.org/10.1016/j.appdev.2013.05.001

Akiva, T., Pearson, L., Sugar, S.A., Peck, S.C., Smith, C., & Denault, A. (2010). *Linking after-school instructional practices to youth engagement: A pattern-centered approach.* Paper presented at the Society for Research on Adolescence Biennial Meeting, Philadelphia.

Aksoy, T., & Link, C.R. (2000). A panel analysis of student mathematics achievement in the us in the 1990s: Does increasing the amount of time in learning activities affect math achievement? *Economics of Education Review, 19*, 2261-2277.

Alexander, K.L., Entwisle, D.R., & Olson, L.S. (2007). Lasting consequences of the summer learning gap. *American Sociological Review, 72*, 167-180.

Allington, R.L., McGill-Franzen, A., Camilli, G., Williams, L., Graff, J., Zeig, J., . . . Nowak, R. (2010). Addressing summer reading setback among economically disadvantaged elementary students. *Reading Psychology, 31*(5), 411-427.

Anderson, L.W., & Krathwohl, D.R. (Eds.). (2001). *A taxonomy for learning, teaching and assessing: A revision of bloom's taxonomy of educational objectives.* New York, NY: Longman.

Arbreton, A., Sheldon, J., Bradshaw, M., Goldsmith, J., Jucovy, L., & Pepper, S. (2008). Advancing achievement: Findings from an independent evaluation of a major after-school initiative *INSIGHT: Lessons learned from the CORAL intiative.* San Francisco: The James Irvine Foundation and Public/Private Ventures.

Arlin, M. (1979). Teacher transitions can disrupt time flow in classrooms. *American Educational Research Journal, 16*(1), 42-56. doi: 10.3102/00028312016001042

Bill and Melinda Gates Foundation. (2010). Working with teachers to develop fair and reliable measures of effective teaching (Vol. June, pp. 1-12).

Blackwell, L.S., Trzesniewski, K.H., & Dweck, C.S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development, 78*(1), 246-263. doi: 10.2307/4139223

Blazevski, J., & Smith, C. (2007). After-school quality and school-day outcomes in michigan's 21st cclc program. Ypsilanti, MI: High/Scope Educational Research Foundation.

Bliese, P. (Ed.). (2000). *Within-group agreement, non-independence, and reliability implications for data aggregation and analysis*: Jossey-Bass.

Bloom, B.S. (1956). *Taxonomy of educational objectives: The classification of educational goals*. New York, NY: Longmans, Green.

Borman, G.D., & Dowling, N.M. (2006). Longitudinal achievement effects of multiyear summer school: Evidence from the teach baltimore randomized field trial. *Educational Evaluation and Policy Analysis, 28*(1), 25-48.

Boss, S., & Railsback, J. (2002). Summer school programs: A look at the research, implications for practice, and program sampler. In E. Cooper Potter (Ed.), *By Request* (pp. 1-85). Washington, D.C.: Northwest Regional Educational Laboratory.

Cameron, C.E., Connor, C.M., & Morrison, F.J. (2005). Effects of variation in teacher organization on classroom functioning. *Journal of School Psychology, 43*(1), 61-85. doi: http://dx.doi.org/10.1016/j.jsp.2004.12.002

Chaplin, D., & Capizzano, J. (2006). Impacts of a summer learning program: A random assignment study of building educated leaders for life: Mathematica Policy Research/The Urban Institute.

Cohen-Vogel, L., Tichnor-Wagner, A., Allen, D., Harrison, C., Kainz, K., Socol, A.R., & Wang, Q. (2014). Implementing educational innovations at scale transforming researchers into continuous improvement scientists. *Educational Policy*, 0895904814560886.

Cooper, H., Charlton, K., Valentine, J.C., & Muhlenbruck, L. (2000). Making the most of summer school: A meta-analytic and narrative review. *Society for Research in Child Development*.

Cooper, H., Nye, B., Charlton, K., Lindsay, J., & Greathouse, S. (1996). The effects of summer vacation on achievement test scores: A narrative and meta-analytic review. *Review of educational research, 66*(3), 227-268.

Cronbach, L.J., Nageswari, R., & Gleser, G.C. (1963). Theory of generalizability: A liberation of reliability theory. *The British Journal of Statistical Psychology, 16*, 137-163.

Denton, C.A., Solari, E.J., Ciancio, D.J., Hecht, S.A., & Swank, P.R. (2010). A pilot study of a kindergarten summer school reading program in high-poverty urban schools. *The Elementary School Journal, 110*(4), 423-439.

Dewey, J. (1938). *Education and experience*. New York: Simon and Schuster.

Diamantopoulos, A., & Siguaw, J.A. (2006). Formative versus reflective indicators in organizational measure development: A comparison and empirical illustration. *British Journal of Management, 17*(4), 263-282.

Duffett, A., Johnson, J., Farkas, S., Kung, S., & Ott, A. (2004). All work and no play: Listening to what kids and parents really want from out-of-school time. In P. Agenda (Ed.): The Wallace Foundation.

Dweck, C. (2007). The perils and promises of praise. *Early Intervention at Every Age, 65*(2), 34-39.

Dweck, C., Walton, G.M., & Cohen, G.L. (2011). Academic tenacity: Mindsets and skills that promote long-term learning (pp. 43). Seattle, WA: Bill and Melinda Gates Foundation.

Eccles, J., & Gootman, J. (Eds.). (2002). *Community programs to promote youth development*. Washington, DC: National Academy Press.

Eccles, J., & Wang, M. (2012). Part 1 commentary: So what is student engagement anyway? In S. L. Christenson (Ed.), *Handbook of research on student engagement*: Springer Science+Business Media.

Farrington, C., Roderick, M., Johnson, D.W., Keyes, T.S., Allensworth, E., Nagaoka, J., & Beechum, N.O. (2012). Teaching adolescents to become learners: The role of noncognitive factors in shaping school performance. *The University of Chicago Consortium on Chicago School Research.*

Fischer, K.W., & Rose, S.P. (1998). Growth cycles of brain and mind. *Educational Leadership, 56*(3), 56-60.

Gordon, E.E., Morgan, R.R., O'Malley, C.J., & Ponticell, J. (2007). *The tutoring revolution: Applying research for best practices, policy implications, and student achievement*. Lanham, MD: Rowman and Littlefield Publishers, Inc.

Graesser, A.C., McNamara, D.S., & VanLehn, K. (2005). Scaffolding deep comprehension strategies through point&query, autotutor, and istart. *Educational Psychologist, 40*(4), 225-234.

Hamre, B., Hatfield, B., Pianta, R., & Jamil, F. (2014). Evidence for general and domain-specific elements of teacher–child interactions: Associations with preschool children's development. *Child Development, 85*(3), 1257-1274.

Harvard Family Research Project. (2004). Promoting quality through professional development: A framework for evaluation (pp. 1-12).

Harvard Family Research Project. (2006). Summer success: Challenges and strategies in creating quality academically focused summer programs. *Issues and Opportunities in Out-of-School Time Evaluation* (Vol. 9). Cambridge, MA: HFRP.

Hattie, J., Biggs, J., & Purdie, N. (1996). Effects of learning skills interventions on student learning: A meta-analysis. *Review of educational research, 66*(2), 99-136. doi: 10.2307/1170605

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research, 77*(1), 81-112.

Hiebert, J., Gallimore, R., & Stigler, J.W. (2002). A knowledge base for the teaching profession: What would it look like and how can we get one? *Educational Researcher, 31*(5), 3.

Jones, S., & Bouffard, S. (2012). Social and emotional learning in schools: From programs to strategies *Social Policy Report* (Vol. 26).

Kim, J. (2004). Summer reading and the ethnic achievement gap. *Journal of Education for Students Placed at Risk, 9*(2), 169-188.

Kim, J.S., & Quinn, D.M. (2013). The effects of summer reading on low-income hildren's literacy achievement from kindergarten to grade 8: A meta-analysis of classroom and home interventions. *Review of educational research, 83*(3), 386-431.

Kreider, H., Caspe, M., Kennedy, S., & Weiss, H. (2007). Family involvement makes a difference. *Family Involvement in Middle School and High School Students' Education,* (3), 112. http://www.hfrp.org/family-involvement/publications-resources/family-involvement-in-middle-and-high-school-students-education Office of Head Start. 2012

Larson, R. (2000). Toward a psychology of positive youth development. *American Psychologist, 55*(5), 170-183.

Lauer, P.A., Akiba, M., Wilkerson, S.B., Apthorp, H.S., Snow, D., & Martin-Glenn, M.L. (2006). Out-of-school-time programs: A meta-analysis of effects for at-risk students. *Review of educational research, 76*(2), 275-313.

Maclellan, E. (2005). Academic achievement: The role of praise in motivating students. *Active Learning in Higher Education, 6*(3), 194-206.

Maslow, A.H. (1943). A theory of human motivation. *Psychological Review, 50*, 370-396.

McCombs, J.S., Augustine, C.H., & Schwartz, H.L. (2011). *Making summer count: How summer programs can boost children's learning*. Santa Monica, CA: RAND Corporation.

McCombs, J.S., Pane, J.F., Augustine, C.H., Schwartz, H.L., Martorell, P., & Zakaras, L. (2014). Ready for fall. Santa Monica, CA: RAND.

McIntosh, K., Herman, K., Sandford, A., McGraw, K., & Florence, K. (2004). Teaching transitions: Techniques for promoting success *between* lessons. *TEACHING Exceptional Children, 37*(1), 32-38.

McTague, B., & Abrams, B. (2011). Access to books: A scaffolded program creates readers. *Reading Improvement, 48*(1), 3-13.

Miller, B.M. (2003). Critical hours: Afterschool programs and educational success: Nellie Mae Education Foundation.

Naftzger, N. (2014). *A summary of three studies exploring the relationship between afterschool program quality and youth outcomes*. Paper presented at the Ready by 21 National Meeting, Covington, KY.

Naftzger, N., Manzeske, D., Nistler, M., Swanlund, A., Rapaport, A., Shields, J., . . . Sugar, S. (2013). Texas 21st century community learning centers: Final evaluation report (pp. 153). Naperville, IL: American Institutes for Research.

Naftzger, N., Tanyu, M., & Stonehill, R. (2010). The impact of self-assessment and quality advisor support on afterschool program quality: Summary of year three findings from wascip quality advisor study. In L. P. Associates (Ed.): William T. Grant Foundation.

Naftzger, N., Vinson, M., Manzeske, D., & Gibbs, C. (2011). New jersey 21st century community learning centers (21st cclc) impact report 2009-2010. In A. I. f. Research (Ed.): New Jersey Department of Education.

Newhouse, C., Neely, P., Freese, J., Lo, J., & Saili, W. (n.d.). Summer matters: How summer learning strengthens student's success: Public Profit.

Pellegrino, J.W.E., & Hilton, M.L.E. (2012). *Education for life and work: Developing transferable knowledge and skills in the 21st century*: National Academies Press.

Piaget, J. (Ed.). (1970). *Piaget's theory*. New York: Wiley.

Pierce, K., Bolt, D., & Vandell, D. (2010). Specific features of after-school program quality: Associations with children's functioning in middle childhood. *American Journal Community Psychology, 45*(3-4), 381-393.

Ramaswamy, R., Gersh, A., Sniegowski, S., McGovern, G., & Smith, C. (2014). Summer learning program quality assessment 2013 phase i pilot report (pp. 27): David P. Weikart Center for Youth Program Quality.

Raudenbush, S., & Sampson, R. (1999). Assessing direct and indirect effects in multilevel designs with latent variables. *Sociological Methods & Research, 28*(2), 123-153.

Roderick, M., Engel, M., & Nagaoka, J. (2003). *Ending social promotion: Results from summer bridge*. Chicago, IL: Consortium on Chicago School Research.

Roseth, C.J., Johnson, D.W., & Johnson, R.T. (2008). Promoting early adolescents' achievement and peer relationships: The effects of cooperative, competitive, and individualistic goal structures. *Psychological Bulletin, 134*(2), 223-246.

Seidman, E. (in press). An emerging action science of social settings. *American Journal Community Psychology*.

Skinner, E.A., & Pitzer, J.R. (2012). Developmental dynamics of student engagement, coping, and everyday resilience (pp. 21-44). Boston, MA: Springer US.

Smith, C. (2013). Moving the needle on "moving the needle": Next stage technical guidance for performance based accountability systems in the expanded learning field with a focus on performance levels for the quality of instructional services. Ypsilanti, MI: David P. Weikart Center for Youth Program Quality, a division of the Forum for Youth Investment.

Smith, C., & Akiva, T. (2008). Quality accountability: Improving fidelity of broad developmentally focused interventions. In H. Yoshikawa & B. Shinn (Eds.), *Transforming social settings: Towards positive youth development*: Oxford University Press.

Smith, C., Akiva, T., Sugar, S., Lo, Y.J., Frank, K.A., Peck, S.C., & Cortina, K.S. (2012). Continuous quality improvement in afterschool settings: Impact findings from the youth program quality intervention study. Ypsilanti, MI: Forum for Youth Investment.

Smith, C., Hallman, S., Hillaker, B., Sugar, S., McGovern, G., & Devaney, E. (2012). Development and early validation evidence for an observational measure of high quality instructional practice for

science, technology, engineering and mathematics in out-of-school time settings: The stem supplement to the youth program quality assessment (pp. 1-25). Ypsilanti, MI: The David P. Weikart Center for Youth Program Quality, a division of the Forum for Youth Investment and Providence Afterschool Alliance.

Smith, C., & Hohmann, C. (2005). Full findings from the youth pqa validation study *High/Scope Youth PQA Technical Report*. Ypsilanti, MI: High/Scope Educational Research Foundation.

Spielberger, J., & Halpern, R. (2002). The role of after-school programs in children's literacy development. Chicago, IL: Chapin Hall Center for Children at the University of Chicago.

Van Egeren, L.A., Bates, L., Sook Lee, K., & Sturdevant-Reed, C. (2011). 21st century community learning centers summer expansion grant: Final evaluation report (pp. 1-36). East Lansing, MI: Community Evaluation and Research Collaborative, University Outreach & Engagement, Michigan State University.

Vygotsky, L. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.

Wheeler, K.A., & Proche, M. (2011). Evaluation results for the summer literacy and learning promotion initiative.

Whitehurst, G.J., Chingos, M.M., & Lindquist, K.M. (2014). Evaluating teachers with classroom observations. Brown Center on Education Policy: Brookings Institute.

Yeager, D., Bryk, A., Muhich, J., Hasuman, H., & Morales, L. (2013). Practical measurement: Carnegie Foundation for the Advancement of Teaching.

Appendix A – Summer Learning PQA Measures

This appendix provides descriptive information regarding the Phase II study sample at all levels of measurement for all Summer Learning PQA measures.

The Summer Learning PQA Form A consists of 74 items nested within18 scales nested within eight domains (Safety, Supportive Environment, Interaction, Planning-Choice-Reflection, Learning Strategies, Higher Order Thinking, Math and Literacy). Table A-1 provides descriptive information for 32 programs at the items, scales and domains as well as the Instructional Total Score (average across all domain scores except Safety, Math and Literacy) and overall Total Score (average across all eight domain scores).

Assessors also completed a checklist related to basic best practices for three transition periods in the program day, greetings, transitions and departures. The Summer Learning PQA Transitions Checklists include 16 items nested within three indexes. Table A-2 provides descriptive information for the 32 programs at the item and index level.

The Summer Learning PQA Form B consists of 13 items nested within four domains (Planning, Staff Training, Family Connections, and Individualization). Table A-3 provides descriptive information for the 32 participating programs at the item and domain levels and an overall Total Score (average across all four domain scores).

Table A-1
Descriptive Statistics for Summer Learning PQA Form A, N= 44 Ratings

	Range	Mean	SD
Safe Environment	2.88	3.63	.72
Psychological and emotional safety is promoted.	3.00	4.36	.84
Positive emotional climate	2.00	4.45	.90
Lack of bias	4.00	4.27	1.37
Removal of Exclusive Behavior	4.00	4.45	1.32
Healthy Environment: The physical environment is safe and free of health hazards.	2.67	4.61	.69
Free of health and safety hazards	4.00	4.64	.89
Clean and sanitary	5.00	4.43	1.17
Suitable for all activities	2.00	4.68	.74

Table A-1 (continued)
Descriptive Statistics for Summer Learning PQA Form A, N= 44 Ratings

	Range	Mean	SD
Emergency Procedures: Appropriate emergency procedures and supplies are present.	3.33	3.65	.92
Posted emergency procedures	5.00	3.75	1.78
Fire extinguisher	5.00	3.20	1.82
First-aid kit	4.00	3.23	1.44
Other safety equipment	5.00	.70	1.68
Supervised entrances	5.00	3.27	1.92
Supervised access to outdoor space	5.00	2.73	2.35
Health and Nutrition: Healthy food and physical activity are provided.	3.33	3.95	.85
Available drinking water	4.00	4.09	1.33
Plentiful food and drinks	5.00	3.70	2.15
Nutritious food and drink	5.00	3.05	2.16
Dedicated physical activity	5.00	3.09	1.85
Supportive Environment	3.60	3.28	.97
Warm Welcome: Staff provides a welcoming atmosphere.	2.67	4.24	.88
Youth Greeted	5.00	2.93	1.96
Staff warm and respectful	2.00	4.55	.85
Positive staff body language	5.00	4.57	1.02
Program Flow: Session flow is planned, presented and paced for youth.	3.20	4.10	.79
Sufficient materials	5.00	4.30	1.21
Explains activities clearly	4.00	4.05	1.18
Appropriate time for activities	4.00	4.14	1.46
Multiple types of activities	4.00	3.86	1.25
Consistent routines and guidelines	4.00	4.05	1.33
Active Learning: Activities support active engagement.	4.00	3.85	1.05
Youth engage with materials or ideas	4.00	4.36	1.28
Youth talk about activities	4.00	3.45	1.66
Balance of concrete and abstract	4.00	4.05	1.40
Tangible products or performances	4.00	3.55	1.80

Table A-1 (continued)

Descriptive Statistics for Summer Learning PQA Form A, N= 44 Ratings

	Range	Mean	SD
Skill Building and Encouragement: Staff encourages and supports youth in building skills.	3.67	3.52	1.14
Learning focus link to activity	5.00	3.20	1.92
Staff encourages youth to try new skills	5.00	3.57	1.63
Staff model skills	4.00	3.45	1.61
Staff breaks down tasks	4.00	3.55	1.74
Staff monitors difficulty	4.00	3.86	1.52
Staff guide initiative in learning	5.00	3.20	1.66
Reframing Conflict: The staff uses youth-centered approaches to reframe conflict.	5.00	.25	.97
Staff approaches calmly	5.00	.30	1.13
Staff seeks youth input	5.00	.25	1.06
Youth examine actions and consequences	5.00	.20	.88
Staff acknowledges and follows up	5.00	.25	1.06
Managing Feelings: The staff encourages children to manage feelings and resolve conflicts appropriately.	5.00	1.82	1.78
Staff acknowledges feelings	5.00	.39	1.06
SA Staff asks children to explain situation	5.00	.57	1.50
SA Helps children respond appropriately	5.00	.52	1.39
SA Children suggest solutions	5.00	.34	.99
Interaction	5.00	1.76	1.72
Belonging: Youth have opportunities to develop a sense of belonging.	5.00	2.88	1.17
Opportunities for children to get to know each other	5.00	3.52	1.30
Values communicated and integrated	5.00	2.84	1.88
Collaboration and Leadership: Youth have opportunity to collaborate and work cooperatively with others.	5.00	2.99	1.15
Interdependent roles	5.00	2.07	1.48
Practice group process skills	5.00	3.20	1.66
Opportunities to demonstrate, explain	5.00	2.93	1.50
All youth lead group	3.00	1.52	.93

Table A-1 (continued)
Descriptive Statistics for Summer Learning PQA Form A, N= 44 Ratings

	Range	Mean	SD
Adult Partners: Youth have opportunities to partner with adults.	5.00	3.14	1.19
Staff shares control with youth	5.00	2.95	1.74
Staff actively involved with youth	5.00	4.52	1.05
Staff and youth accountable to expectations	5.00	2.57	2.11
Positive behavior management style	5.00	3.02	2.11
Engagement	3.17	3.46	.86
Planning, Choice, and Reflection: Youth have opportunity to direct their own learning.	5.00	2.51	1.07
Opportunities to make plans	5.00	2.30	1.64
Content alternatives	5.00	2.43	1.61
Process alternatives	5.00	2.80	1.80
Intentional reflection	5.00	2.52	1.85
Structured opportunities to provide feedback	5.00	2.30	1.80
Learning how to learn: Youth are supported developing learning initiative and persistence.	4.00	2.92	1.10
Problem-solve for improvement	5.00	3.05	1.74
Identify learning strategies	4.00	1.86	1.32
Effort-achievement beliefs	5.00	3.50	1.66
Higher Order Thinking: Youth are supported in developing higher order thinking skills.	5.00	2.08	1.88
Staff encourages youth to deepen knowledge	5.00	2.57	1.73
Connecting activity and other knowledge	5.00	3.34	1.83
Encourage use of creativity, curiosity, or imagination	5.00	3.16	1.70
Total Score	1.85	3.52	.51
Instructional Total Score	2.01	3.31	.61
Math: Youth are supported in mathematical problem solving.	5.00	1.62	1.91
Participate in problem solving	5.00	1.50	2.19
Opportunities to apply knowledge and skills	5.00	1.36	1.98
Use reasoning to evaluate	5.00	.95	1.72
Linking concrete examples	5.00	.98	1.72
Support the conveying of concepts	5.00	.84	1.51

Table A-1 (continued)
Descriptive Statistics for Summer Learning PQA Form A, N= 44 Ratings

	Range	Mean	SD
Literacy: Youth are supported in reading and writing.	5.00	3.33	1.70
Participate in literacy activities	5.00	2.32	2.27
Opportunities to read in multiple settings	5.00	1.95	2.08
Staff encourage expression in writing	5.00	1.64	2.05
Vocabulary discussed	5.00	1.89	2.18
Available materials and reading environment	5.00	1.89	2.14
Multiple reading and writing activities	5.00	2.07	2.22

Table A-2
Descriptive Statistics for Summer Learning PQA Transition Checklists, N= 44 Ratings

	Range	Mean	SD
Greetings - Opening and arrival time	1	0.59	0.30
Children greeted by staff	1	0.76	0.43
Session starts within 10 minutes of scheduled time	1	0.8	0.41
Welcoming activity or icebreaker	1	0.39	0.49
Incorporates themes or aspects of program culture	1	0.43	0.50
Transitions: Group moves to new activity	1	0.90	0.24
Smooth and quick transition times	1	0.89	0.32
Clear transition communication	1	0.94	0.24
On task and ready for transition	1	0.88	0.33
Activity choices clearly communicated	1	0.86	0.35
Program lessons incorporated	0	1	0
Departure: When children leave for the day	0	1	0
Organized process	0	1	0
Smooth process	0	1	0
Constructive activities while waiting	0	1	0
Children left unattended	0	1	0
Utilizes parent engagement opportunity	0	1	0
Verification system	0	1	0
Program incorporated	0	1	0

Table A-3

Descriptive Statistics for Summer Learning PQA Form B, N= 32 Interviews

| Scale/Item	Range	Mean	SD
Organizational planning	4	2.89	1.00
Proactive Planning	4	2.69	1.36
Youth Input	4	2.50	1.61
Lesson Plan Framework	4	3.56	1.63
Staff Training	3.5	3.27	0.95
Staff Training	4	3.69	1.31
Support for Non-certified Teachers	4	3.39	1.67
Staff Meetings	4	3.56	1.46
Staff Observation Feedback	4	2.44	1.70
Family Connections	4	3.09	1.25
Year-Round Contact with Families	4	3.31	1.69
Relationship-building with Families	4	2.87	1.52
Individualization	4	3.32	1.58
Youth Assessment	4	3.71	1.83
Individualized, Tailored Instruction	4	2.94	1.75
Interview total average score	3.71	3.09	0.87

Appendix B - Logic Model and Review of Literature

Logic Model

The quality instructional practices and the quality management practices described above fit in a logic model shown in Figure B-1. This logic model (Smith, 2013) describes relationships between summer learning service outputs – staff practices, staff and youth engagement, in-program skill – and ultimately desired school success outcomes. This logic model highlights the need to specify the links between specific roles (i.e., managers, instructors, youth) and effective practices at a fairly granular level. This level of granularity is necessary – aligning specific roles with specific behaviors or practices – because the goal of the Summer Learning Program Quality Intervention (SLPQI) is to produce effective performance feedback. Figure B-1 highlights several additional points that have and will continue to inform our work:

- Student engagement during instruction is a critical contributor to learning. Numerous studies have documented the relationship between engagement during instruction and school success outcomes (Eccles & Wang, 2012; Skinner & Pitzer, 2012).

- Many summer learning programs focus on child and youth development broadly conceived, including social and emotional or 21st Century skills. These skills are likely demonstrated during summer learning programs and may be important targets for evaluation of program effectiveness (Van Egeren, Bates, Sook Lee, & Sturdevant-Reed, 2011).

- The demonstration of a skill that was developed in a summer learning program but is demonstrated in another setting such as a school day classroom likely requires attention to issues of learning transfer (Pellegrino & Hilton, 2012).

Figure B-1
Service Production Map for Summer Programs
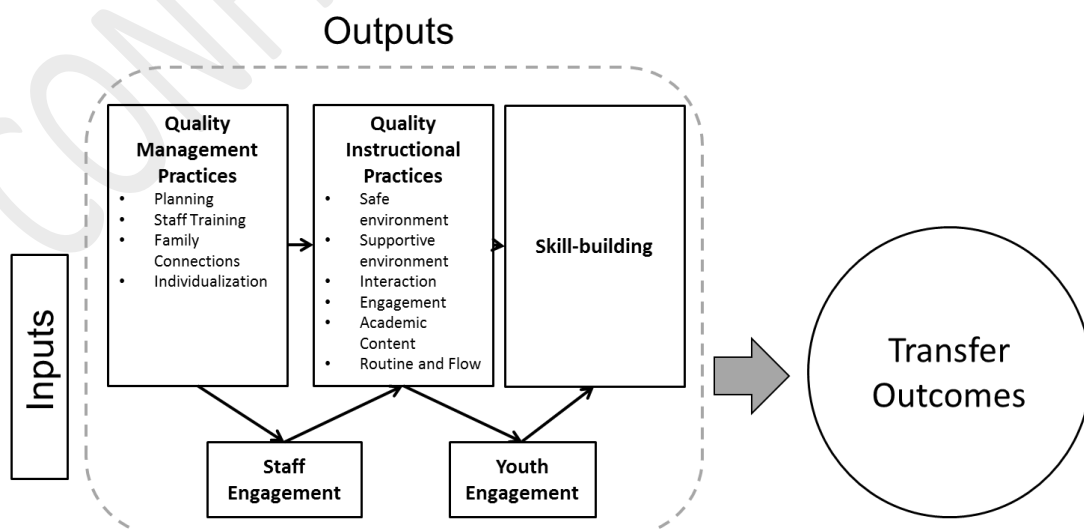
Review of Literature

Summer learning programs are unique environments where the positive youth development focus is equal parts engagement and academic skill-building. There are two theoretical underpinnings in this positive youth development frame: (1) engagement stems from the intersection of youth's sense of belonging and interest and the right level of challenge of the activity (Akiva, Cortina, Eccles, & Smith, 2013) and (2) academic and other skill-building content rely on specific teaching and learning strategies (Dweck, 2007; Hattie & Timperley, 2007; Maclellan, 2005). These theories overlap in a typical summer learning program day, where certain times of day may be designed for different purposes – enrichment or academic skill-building. The Summer Learning PQA is designed to measure both of these purposes and detect practices within each frame.

The scales and items of the Summer Learning PQA are intended to assess practices that support quality summer programming, specifically programming that promotes learning and stems summer learning loss. The practices assessed were drawn from a literature review of summer programming research and research on quality instructional practices in out-of-school time programs and research and theory on youth development and learning. This abridged literature review will outline literature support and rationale for the quality instructional practices of the domains and scales included in Form A of the Summer Learning PQA, the quality management practices assessed through interviews (Form B) and describe the relationship of these broad domains to the logic model for quality summer programming. Special attention will be given to new scales developed to focus particularly on academic learning and supports.

Quality Instructional Practices

PQA Domains Assessing Positive Youth Environments

High quality summer learning programs must first of all be high quality youth programs. The basic structure and standards of the Youth PQA are present in the Summer Learning PQA as foundational measures of program quality. Evidence is emerging that youth participation in high quality programs, as defined by validated measures such as the Youth PQA, is associated with academic achievement, less grade retention and/or fewer disciplinary referrals (Blazevski & Smith, 2007; Naftzger, 2014). It follows that summer programs that emphasize both youth development or social-emotional growth and academics produce positive outcomes, "higher school-year attendance and achievement, increased motivation to learn, increased feelings of belonging, and reduced participation in risky behavior. These positive outcomes are most likely to result when programs begin in the early grades, are offered over multiple summers, and focus on prevention and development rather than remediation" (Newhouse et al., n.d., p. 8).

With this understanding that effective summer programs employ positive youth development practices as well as focus explicitly on academic skill-building. The following sections will briefly outline literature support for the broad domains and features of positive youth development that guided the revision of the Summer Learning PQA.

Safe Environment. According to Maslow's hierarchy of needs, physical and psychological safety needs form the base from which further growth and exploration can proceed (Maslow, 1943). Physical and psychological safety is one of the features of positive youth development programs specified by the National Research Council (Eccles & Gootman, 2002). Safety practices are typically part of licensing requirements, but basic facets are included in the Summer Learning PQA.

Supportive Environment. A warm, positive, and welcoming environment for youth or children is conducive to learning (Gordon, Morgan, O'Malley, & Ponticell, 2007; Pierce, Bolt, & Vandell, 2010). Positive developmental programs for youth are situated in positive relationships between caring adults and youth participants (Eccles & Gootman, 2002). Key indicators of quality include effective encouragement and feedback, which are essential to supporting the development of academic skills (Dweck, 2007; Hattie & Timperley, 2007; Maclellan, 2005). The Supportive Environment domain of the Youth PQA identifies instructional practices that facilitate active engagement of youth within activities, staff encouragement of youth and intentional conflict management that empowers youth.

Interaction. The adult behaviors associated with a supportive environment provide the base for positive interaction and experiences comprised of collaboration, shared leadership, and a sense of belonging. A meta-analysis of 148 studies concluded cooperative learning boosts achievement and peer relationships (Roseth, Johnson, & Johnson, 2008). Research also shows that academic success is linked to youth feeling a sense of belonging in a learning community and believing that teachers or other adults care about their well-being. A critical research review puts a sense of belonging as one of the mindsets that set the stage for improved learning outcomes by affecting academic perseverance, behaviors, and learning strategies (Farrington et al., 2012). The Interaction domain of the Youth PQA describes how staff support youth engagement and collaboration among youth. Items in this domain include the establishment of collaborative learning groups and leadership opportunities for all youth.

PQA Scales and Domains with Academic Learning Focus

Engagement. The final instructional domain of the Summer Learning PQA is comprised of three scales: Planning, Choice, Reflection; Learning How to Learn; and Higher Order Thinking. This domain builds on the Supportive Environment domain, positing that for lasting learning to occur, youth must be actively engaged in the process, a tenet of constructivist theories of learning. These theories indicate that in order for learning to be retained well and transfer to new settings it must be thoroughly practiced and

integrated with existing knowledge schemas. Engaged learning involves youth learning how to learn and employing learning strategies such as planning and reflection that enable them to transfer and apply learning in different contexts (Gordon et al., 2007; Hattie, Biggs, & Purdie, 1996). Learning environments that are supportive, interactive and engaging promote interest, challenge and belonging which have been associated with higher levels of youth engagement (Akiva et al, 2011). The Summer Learning PQA also includes four additional scales not included in the Youth PQA: Learning How to Learn and Higher Order Thinking which stress the importance of helping youth to develop and apply learning strategies and specific scales for the content areas of math and literacy.

Learning How to Learn. The Learning How to Learn items incorporates major elements identified in the Consortium on Chicago School Research's model for teaching adolescents to become learners: academic mindsets, learning strategies and academic perseverance (Farrington et al., 2012). Most students have not learned the metacognition skills necessary for deeper learning and the lack of emphasis on teaching those skills has become a growing concern (Graesser, McNamara, & VanLehn, 2005). Similarly, based on a review of the literature, Yeager and colleagues include "skills, habits, and know-how" and "believing that they are capable of learning…" as key elements of a practical theory of productive persistence (Yeager, Bryk, Muhich, Hasuman, & Morales, 2013). In addition, research by Carol Dweck and others has highlighted the importance of academic tenacity and children understanding that mistakes or errors are not an indication of lack of intelligence or a limitation of the ability to learn (Blackwell, Trzesniewski, & Dweck, 2007; Dweck, Walton, & Cohen, 2011).

Higher Order Thinking. This scale supports learning that is retained well and transfers to new settings because it is integrated and linked with existing knowledge schemas as well as thoroughly practiced with conscious learning strategies (Gordon et al., 2007; Hattie et al., 1996). In line with prominent constructivist learning theories such as those of Dewey (1938), Piaget (Piaget, 1970), Vygotsky. (1978) Bloom (1956), Fischer (Fischer & Rose, 1998) and others, the PQA sees youth as actively constructing knowledge and as participants in their learning and not passive recipients of knowledge that is dispensed by adult experts. Youth go through a progression of learning towards more integration and complexity, as represented by Benjamin Bloom's (1956) taxonomy. Higher order thinking refers to upper categories of Bloom's taxonomy. The taxonomy, revised slightly by psychologists in the 1990's, is widely used to generate questions that move students from simply remembering facts to increasingly complex forms of thinking (Anderson & Krathwohl, 2001). The items in this scale assess whether staff support youth in advancing through Bloom's taxonomy. Additionally, linking content and session activities to youth's personal interests enhances motivation (Akiva, 2012).

Academic Content Area. In order to connect the measures directly to target areas for summer learning, two scales tied to best practices in specific content areas were included in the revised measures. These two additional scales emphasize practices that support learning in literacy and math:

- *Literacy* - Summer Learning programs have great potential to at least prevent summer skill loss and even promote gains (Borman & Dowling, 2006; Chaplin & Capizzano, 2006; Harris Cooper et al., 1996; Denton, Solari, Ciancio, Hecht, & Swank, 2010; J. S. Kim & Quinn, 2013; Wheeler & Proche, 2011). For economically disadvantaged youth, merely providing access to books and spending more time reading has significant effects (Allington et al., 2010; J. Kim, 2004; McTague & Abrams, 2011). With this potential in mind, the Literacy scale of the Summer Learning PQA emphasizes providing reading opportunities and access to reading materials in multiple contexts throughout the program day.

- *Math* - Summer learning programs can have small, but statistically significant effects on math achievement (Lauer et al., 2006). Time spent on math practice, in or out of school, has a proportionate effect on math achievement (Aksoy & Link, 2000). The National Council of Teachers of Mathematics, in Principles and Standards for School Mathematics (2003), proposed a vision for mathematics teaching that includes problem-solving, reasoning, connections, representations along with number and operations, and geometry and measurement. The Common Core Standards (2013) for Mathematical Practice include making sense of problems, reasoning abstractly, constructing arguments, modeling and the use of tools.

Routine and Flow Checklists. Transitions can be moments where time is wasted and misbehavior is more likely, especially for some children with developmental challenges (Arlin, 1979; Cameron, Connor, & Morrison, 2005; McIntosh, Herman, Sandford, McGraw, & Florence, 2004). Since transitions are part of the program environment, the transitions checklist assesses whether these shifts in schedule are smooth and organized. Departure Time and Greeting Time are specific transitions that have their own checklists. Greeting time can promote welcome and orientation to the day, while departure is a transition with particular opportunities and concerns such as intersection with family members and safety issues.

Quality Management Practices

The Summer Learning PQA Site Lead Interview was constructed around four domains to highlight key practices in the structure of a high quality summer program. "Quality" is well-defined and rooted in research. A major study from RAND shows that individualized academic instruction, parental involvement and smaller class sizes are components of high quality programs that are making a difference (McCombs et al., 2011).

Planning. Summer programs, especially school-based summer programs, struggle to plan in advance and plan collaboratively for quality programming (Boss & Railsback, 2002). However, research and best practice examples from the field indicate that proactive planning is necessary for high quality summer programming. There is a great opportunity to systemically improve program quality through valid and reliable indicators of program length, planning cycle and other time/resource inputs. Literature has pointed to several key components, including youth input, that form the foundation for quality planning indicators in afterschool and in summer (Miller, 2003).

Staff Training. Staffing is a key component of quality in out-of-school time (Harvard Family Research Project, 2004). A large body of research on characteristics of high quality summer programs has shown that qualified, caring staff have a great impact on quality at the point-of-service. Many staff who work directly with youth lack pre-service training, specific credentials or degrees. While the identity of the out-of-school-time worker may still be evolving, the significance of professional development is emerging as an important driver of program quality and positive social and cognitive outcomes for youth (Harvard Family Research Project, 2004). Staff should be qualified and equipped to take responsibility for managing the program's day-to-day operations. Programs identified by NSLA that empower their staff to make decisions and grow professionally are more likely to achieve sustainable quality. Overarching "professional development systems," or combinations of various professional development modalities and providers, such as those described in NSLA standards, have shown the largest sustained effects on program quality when evaluated (Harvard Family Research Project, 2004).

Family Connections. Partnerships between programs and families can provide academically focused summer programs with numerous advantages. Linkages between families and out-of-school-time programs can also improve family involvement in children's education and family relationships more broadly and may also improve program implementation outcomes (Harvard Family Research Project, 2006). Further research notes that there are several program components related to improved achievement for summer program attendees including parent involvement and participation (Harris. Cooper, Charlton, Valentine, & Muhlenbruck, 2000).

Individualization. Youth in summer programs often enter with a variety of academic and developmental needs, which makes it essential for programs to develop a standardized youth assessment

method, aligned with program goals, using formal and informal measures. High quality summer programs incorporate youth assessments seamlessly into their program design in order to track and report on youth outcomes. Pre-program assessments should be aligned with program goals and should inform instructional strategies and differentiation during the session. In addition, pre-program assessments should serve as a baseline measure when assessing achievement of goals at the end the program. Research has found that small-group or individualized instruction is related to improved achievement for summer programs (Harris. Cooper et al., 2000).

Appendix C - Reliability and Validity of Summer Learning PQA Data

Evaluating reliability and validity of data from observation-based measures of settings requires cautious application of standard psychometric concepts and tools (Cronbach, Nageswari, & Gleser, 1963; Raudenbush & Sampson, 1999; Seidman, in press) and careful alignment between (a) the different purposes for which scores will be used and (b) the different methods to determine score reliability and validity. Specific challenges include:

- The instructional practices recommended by experts may not occur in all settings all the time. Observational measures and methods of data collection that are not calibrated to offering structure and sequence may both miss critical practices that do in fact occur or, produce low scores for practices which are not part of the curriculum.

- Many setting-level measurement constructs are formative rather than reflective in nature, meaning that the items grouped within a given scale may not "reflect" a construct that exists independently of the items. Formative constructs do not necessarily exhibit "internal consistency" among items and are often better understood as indexes.

- Facets of data collection – items, raters, time of day and year, programs, and interactions of these facets - may introduce substantial error into quality ratings. These sources of unreliability can only be detected with complex data collection designs that "cross" raters and other important facets of observational measurement so that score variance may be partitioned.

There is often pressure to improve score reliability, even when at cross-purposes with more important goals for validity. For example, a single total score with high internal consistency, high construct validity and low rater bias may be achieved by deleting many items from the PQA and may serve purposes of differentiating between high and low quality sites. However, for learning and behavior change purposes less reliable scores that describe specific staff behaviors or sets of practices that typically co-occur may be more useful.

For these reasons our approach to the assessment of the reliability and validity of Summer Learning PQA consisted of a set of steps, following the Weikart Center's approach to the development of observational measures (Smith, Hallman, et al., 2012), which were designed to maximize our understanding of these complex issues within the limitations imposed by the project budget. The final sample used for these reliability and validity analyses includes 44 unique AM and PM session ratings collected at 32 programs sites by 18 assessors.

Step 1 – Content and substantive validity

As a part of the item development process (and before data was collected), we examined two forms of validity for the new academic quality items - which are those presented in Table C-1 – using feedback from 44 expert practitioners (a mix of assessors and site coordinators from Phases I and II). We use the term *content validity* to denote the perception by summer learning program experts that a practice named in the Summer Learning PQA is important in their practice, i.e., that it is worth taking the time and trouble to measure. In general, experts agreed that the new items were important to academically oriented program settings with an average rating across all new items of 2.89 on a three point scale where 3 = important and 1 = not important. The lowest average importance score for any single item was 2.55. These findings suggest that expert practitioners endorse the content of the new academically focused item as being important to high quality summer learning practice.

We use the term *substantive validity* to indicate that the items describe practices that occur with a high enough prevalence in daily sessions that they are likely to be observed on any given day. If the practice were important but rarely observed, then a single observation of the program is likely to miss that practice, suggesting that the item does not adequately sample from a domain of practices being observed. In this case, we asked our expert raters to tell us how likely each of the new Summer Learning PQA items were to be observed during a typical academically oriented session where 1 = not likely, less than 40 percent of the time; 2 = somewhat likely, 40-79 percent of the time; 3 = quite likely, 80 percent of the time or more. In this case, we might characterize our findings for substantive validity as moderate, although no real benchmark exists for this judgment. Six of the 17 items scored below 2 on the three point scale, indicating that these practices could be seen in less than 40 percent of sessions. Several of the items were scored quite low in the Learning Strategies and Math scales, indicating that the validity of these items may be lower because, while important practices, the instructional model being deployed during any given session does not always include them.

It is important to remember that this argument can also be turned around: The instructional model being deployed is invalid because it does not include these critical practices. For this reason it is important that programs and systems using observational ratings should select only the items that they believe are appropriate to their instructional model.

Table C-1

Expert Practitioner Importance and Prevalence in a Typical High Quality Academic Session, N=44

	Average Importance	Average Prevalence
Learning Strategies		
Problem-solve for improvement	2.91	2.18
Identify learning strategies	2.55	1.64
Effort-achievement beliefs	2.82	1.82
Higher Order Thinking		
Staff encourages youth to deepen knowledge	2.91	2.09
Connecting activity and other knowledge	3.00	2.09
Encourage use of creativity, curiosity, or imagination	2.73	2.09
Math		
Participate in problem solving	2.80	2.10
Opportunities to apply knowledge and skills	2.60	1.80
Use reasoning to evaluate	2.70	1.60
Linking concrete examples	2.60	1.90
Support the conveying of concepts	2.60	1.80
Literacy		
Participate in literacy activities	2.91	2.64
Opportunities to read in multiple settings	2.91	2.55
Staff encourage expression in writing	3.00	2.18
Vocabulary discussed	2.64	2.18
Available materials and reading environment	2.73	2.00
Multiple reading and writing activities	2.82	2.36

Step 2 –Reliability

Table C-2 describes the distribution of Summer Learning PQA scores for the domains and scales that constitute the Instructional Total Score on the Summer Learning PQA. The Instructional Total Score is comprised of all Summer Learning PQA measures but with the Safety domain removed.[13] As the name suggests, the Instructional Total Score includes items which are under the control of the instructor and likely to vary as a function of the instructor's skills and the curriculum model being deployed. Table C-2 suggests that quality scores are distributed across the entire length of the scale (range and mean), with substantial variance across ratings (standard deviation), and under a fairly normal distribution (skew statistics between -2 and +2). These characteristics of the data support further analyses described in this

---

[13] Because the safety domain is a global measure of environmental factors that are frequently governed by policy (e.g., licensing) there is frequently little variance on this scale.

appendix and suggest that the SLPQI scale does not present severe floor or ceiling effects which could hamper efforts to detect change over time in quality ratings.

Table C-2
Descriptive Statistics and Alpha-Reliability for All Domains/Scales of Summer Learning PQA

	Range	Mean	SD	Skew	Scale α
Supportive Environment	3.60	3.28	.97	.15	.76
Warm Welcome	2.67	4.24	.88	-.82	.27
Program Flow	3.20	4.10	.79	-1.03	.60
Active Learning	4.00	3.85	1.05	-1.12	.61
Skill Building and Encouragement	3.67	3.52	1.14	-.24	.76
Reframing Conflict	5.00	.25	.97	NA	.95
Managing Feelings	5.00	1.82	1.78	.32	.96
Interaction	5.00	1.76	1.72	.14	.59
Belonging	5.00	2.88	1.17	.14	.40
Collaboration and Leadership	5.00	2.99	1.15	-.19	.39
Adult Partners	5.00	3.14	1.19	-.37	.35
Engagement	3.17	3.46	.86	-.61	.71
Planning, Choice, and Reflection	5.00	2.51	1.07	.31	.65
Learning How to Learn	4.00	2.92	1.10	.12	.19
Higher Order Thinking	5.00	2.08	1.88	.22	.65
Instructional Total Score	2.01	3.31	.61	.03	.87
Math	5.00	1.62	1.91	.62	.94
Literacy	5.00	3.33	1.70	-1.17	.94

Table C-2 also provides alpha coefficients for Summer Learning PQA domains, scales and the Instructional Total Score. For this type of reliability, coefficients of 0.7 or above are preferred. In Table D-2, only nine of 18 coefficients are at 0.7 or above. However, we present the alpha coefficients only for reference. We argue that the SLPQI scales are formative in nature, (Diamantopoulos & Siguaw, 2006) making alpha-type reliability inappropriate. In contrast, we treat the domain scores and Instructional Total Score as more like traditional reflective scales where the alpha coefficient is more appropriately applied. In this case, all but one of the four domain and total scores are in the acceptable range, with total score coefficient at 0.87.

Inter-Rater Reliability

Inter-rater reliability is an important form of reliability for observational measures. Eighteen assessors were selected to complete Summer Learning PQA assessments for the study. This pool of assessors was selected from existing lists of assessors in each city already certified on the PQA. For this study the assessors participated in an additional rater training, either live or distance which require achievement of 80 percent perfect agreement with gold standard scores for a test video. The pool of 18 raters completed 26 tests of reliability for the PQA during the project period and the average percent perfect agreement across 68 test items was 91 percent.

Test-Retest Reliability

The Phase II study did not include data collection at multiple time points. However, another recent sample did include data at two time points on the Learn How to Learn and Higher Order thinking scales. [14] 24 paired observations were made across two time periods, one week apart. Significant correlations were found at the scale level for both Learning How to Learn ($r = .70$) and Higher Order Thinking ($r = .74$) indicating a moderate/high level of consistency across a short time interval.

Reliability of an Aggregate Program Quality Rating

Another important aspect of reliability (consistency) in quality ratings for summer learning programs relates to how much alike the AM and PM offerings were in the same program. This issue is important if we want to combine the AM and PM scores to produce an average for the program. If the scores within each organization are consistent, that provides greater warrant to create an average quality rating for the site.

In order to better understand the consistency of AM and PM quality rating within each program, we calculated two intra-class correlations (ICC (1) and ICC (2)) which are provided in Table C-3. These coefficients describe the reliability of multiple quality ratings from the same program site, in this case the degree of agreement between AM and PM ratings. In general, higher levels of agreement among ratings from the same program site are required to meaningfully interpret an average score for multiple respondents in the same program site. ICC (1) can be understood as the reliability of a rating from a single respondent and the proportion of scale score variance explained by differences between programs. ICC

---

[14] The following analyses were drawn from reliability and validity analyses for Academic Skill-Building Program Quality Assessment, a PQA-type measure developed for use in tutoring programs (Hillaker, Smith, McGovern, and Sniegowski, 2014).

(2) describes the reliability of the scale mean for each program by taking into account the number of additional raters included in the mean scale score (Bliese, 2000). In general, ICCs (1) and (2) indicate that there is relatively high agreement within program sites and that program site means can be meaningfully interpreted, except in the case of the Literacy score which appears to have very little consistency between AM and PM sessions in this sample.

Table C-3
Intraclass Correlations (ICC) for Instructional Total Score, Plan-Choice-Reflect Scale and New Academic Scale Scores, N=22 Programs with Both AM and PM Ratings

	ICC(1)	ICC(2)
Instructional Total Score	.08	.66
Planning, Choice, and Reflection	.66	.80
Learning How to Learn	.51	.68
Higher Order Thinking	.38	.55
Math	.39	.56
Literacy	.01	.02

ICCs (1) and (2) were calculated using variance estimates from one-way ANOVA with random effects model for the data with each scale as the dependent variable and the site ID as the factor. The formulas for each are provided in Figure C-1 where MSB is the scale score variance accounted for between sites, MSW is the scale score variance accounted for within sites and K is the average number of ratings contributing to the mean score for that program.

Figure C-1
Calculating Formulas for Intraclass Coefficients

$$ICC(1) = \frac{MSB - MSW}{MSB + [(k-1)*MSW]} \qquad ICC(2) = \frac{k(ICC(1))}{1 + (k-1)ICC(1)}$$

Step 3 – Convergent Validity

Convergent validity refers to the co-occurrence of higher or lower quality ratings with other factors that we predict in advance to be associated with higher and lower quality. Because the Phase II study was focused on better understanding implementation of the SLPQI over a small number of programs that vary by type, opportunities to collect additional sources of data that could be used for these purposes were limited. However, there are three opportunities to test working hypotheses about

convergent validity related to academic sessions, management and instructional practices, and instructional practices and child engagement. Specifically we predicted that:

- Summer Learning PQA scales focused on the quality of academic instruction should score higher for session which include academic content compared to sessions that do not.
- Programs with higher scores for instructional quality should also have higher scores for management practices. The correlation should be highest between the Form A Instructional Total Score and the Form B Individualization scale score because individualization is an instructional practice.
- Form A ratings should be positively correlated with measures of child engagement.

Academic Content. Because several of the new scales produced for the Summer Learning PQA were specifically designed for use in academically focused program sessions, we hypothesized that academically oriented sessions should score more highly than non-academically oriented sessions. Figures 3 and 4 in part V of this report indicate that for each of the academic quality domains – Learning Strategies, Higher Order Thinking, Math, Literacy – the academically oriented sessions occurring in the morning scored substantially higher than the enrichment oriented sessions in the afternoon.

Management Practices. Table C-4 provides Pearson-r correlation coefficients between the domain and total scores for the Summer Learning PQA Form B and the Form A Instructional Total Score. In general the correlations are substantively large, indicating that quality of management practices and quality of instructional practices tends to be higher or lower in the same programs. The two total scores are associated at r=0.49 while, as predicted, the Individualization score is the most highly correlated of the Form B component scores with the Form A Instructional Total Score with r=0.43.

Table C-4
Correlation Between Management Quality and Instructional Quality, N=32

	Form A Instructional Total Score
Organization Planning	0.40
Staff Training	0.18
Family Connections	0.28
Individualization	0.43
Form B Total Score	0.49

Child Engagement. The Observation of Child Engagement (OCE) was administered with the Summer Learning PQA by a trained assessor in all 44 rated sessions. For each of these offerings, the OCE measure generated between one and 12 ratings reflecting the level of group engagement at 10-minute intervals. The average across these observations was created to produce a single rating of student engagement during each of the 12 program sessions. Table C-5 provides Pearson-r correlations for the Child Engagement score and the Instructional Total Score for AM and PM sessions. Even in this very crude construction of the Child Engagement score, the pattern of correlation is in the expected direction, the highest correlations occurring within the AM and PM sessions and the lowest correlations between morning engagement and afternoon quality or the reverse.

Table C-5
Correlations for Child Engagement and Instructional Quality, AM and PM

	1	2	3	4
1. AM Child Engagement				
2. PM Child Engagement	.68			
3. AM Instructional Quality	.18	.08		
4. PM Instructional Quality	-.10	.26		

Correlation Across The Eight Form A Instructional Quality Domains

As a final step in the discussion of validity of for the Form A measures of instructional quality, Table C-7 presents a correlation matrix for the eight Form A domains. From the table it is evident that the eight domains exhibit associations ranging from -.276 to .739. Perhaps most importantly, the overall pattern reflects modest levels of association (1) that we would expect given a theory that high quality practices should co-occur across domains in the same programs and (2) that also suggest that our measures are not measuring the same things (i.e., they are not too highly correlated). Also of interest is the fairly low level of correlation between the math and literacy domain scores (r=.216) and these scores and the all of the other domain scores, suggesting that these academic scales do not have high overlap with the other measures. Finally, the high correlations between the Supportive Environment domain score and the Learning Strategies and Higher Order Thinking scales is not surprising given that the content of these scales is similar, with the new Learning Strategies and Higher Order Thinking scales operating as deeper extensions of the Active Learning and Skill Building scales in the Supportive Environment domain.

Table C-7
Correlation Across Eight Domains of Instructional Quality

	1	2	3	4	5	6	7
1. Safety							
2.Supp Environ	.60						
3.Interaction	.49	.63					
4.Plan-choice-refl	.05	.36	.05				
5.Learning Strat	.48	.63	.18	.14			
6.High Order Th	.53	.74	.58	.31	.37		
7. Math	.07	.13	-.17	.18	.18	.24	
8. Literacy	-.23	-.11	-.28	.24	.13	-.10	.22

Appendix D - Missing Data

One primary challenge in the Phase II study comes from the very low response rate for the site coordinator survey: a post-study report on implementation by site coordinators. Site coordinators were contacted about the implementation survey on multiple occasions (9/15, 9/24, 10/6) via both e-mail and phone. System leads were also engaged to reach out to program managers (10/15). Our explanation for the low response rate is essentially that we waited too long to post the invitation to complete the survey. Specifically, our decision to wait until all summer programs were completed - the last week in August – allowed program staff to move on to vacations (e.g., not seeing e-mails or not using district email accounts) or new jobs.

Table D-1 describes the pattern of response and non-response in the study sample. Overall post-study reports on implementation were available for five of 12 (42 percent) sites in the Northern California system, nine of 9 (100 percent) sites in the City of Seattle system, one of 4 (25 percent) in the School's Out Washington System, two of 4 in the Seattle Public Schools System, and three of 3 (100 percent) sites in the Grand Rapids system. Overall for the 32 programs included in the study, the total response rate is 63 percent. In addition to post-program information from these 20 programs, we also conducted follow-up interviews with an additional five staff in five of those same programs. The data from which our conclusions about implementation and customer satisfaction were drawn come from a total of 25 unique individuals at 20 programs.

Table D-1

Study Systems and Programs by Completed Survey and Follow-up Interview

Network	Organization	Program	Supervisor Survey Completed	Site Coordinator Follow-up Interview	Missing
Northern Cal	Child Development Centers	Jacobson 21			X
Northern Cal	Hsng Auth Cnty San Joaquin	Kids N Motion - Nevin		X	
Northern Cal	Hsng Auth Cnty San Joaquin	Kids N Motion - SR			
Northern Cal	Stockton Unif School District	Stockton Summer U-Con		X	
Northern Cal	Stockton Unif School District	Stockton Summer U-SV		X	
Northern Cal	Stockton Unif School District	Fremont School	X		
Northern Cal	Stockton Unif School District	Harrison School			X
Northern Cal	Stockton Unif School District	Pittman School			X
Northern Cal	Stockton Unif School District	Roosevelt School			X
Northern Cal	YMCA of San Joaquin Cnty	Lincoln High Sch	X		
Northern Cal	YMCA of San Joaquin Cnty	McKinley Park			X
Northern Cal	Bay Area CommResources	Urban Promise Academy			X
City of Seattle	City of Seattle Prk & Rec	Parks Acd of Learning	X	X	
City of Seattle	City of Seattle Prk & Rec	Eckstein Middle School	X		
City of Seattle	City of Seattle Prk & Rec	Mercer Middle School	X		
City of Seattle	City of Seattle Prk & Rec	Northgate Elem School	X	X	
City of Seattle	Refugee Women's Alliance	Seattle World School	X	X	
City of Seattle	YMCA of Greater Seattle	Chief Stealth High Sch	X		
City of Seattle	YMCA of Greater Seattle	Cleveland High School	X		
City of Seattle	YMCA of Greater Seattle	Franklin High School	X		
City of Seattle	YMCA of Greater Seattle	Madrona Elementary Sch	X		
SOW – Feed YB	America SCORES	P.A.S.S. at Maple	X		
SOW – Feed YB	Somali Comm Svcs Coalition	Tukwila			X
SOW – Feed YB	SW Youth & Family Services	New Futures - Sum Learn			X
SOW – Feed YB	White Center Heights Elem	Highline - Bridges			X
Seattle PS	Seattle Public Schools	Hawthorne Elementary		X	X
Seattle PS	Seattle Public Schools	Highland Park Elem			X
Seattle PS	Seattle Public Schools	Martin LuthKing Jr. Elem	X		
Seattle PS	Seattle Public Schools	Viewlands Elementary			X
Grand Rapids	Campfire	Burton Middle School	X	X	
Grand Rapids	Campfire	Burton MS - MLK	X	X	
Grand Rapids	Cook Library Scholars	Cook Library Scholars		X	