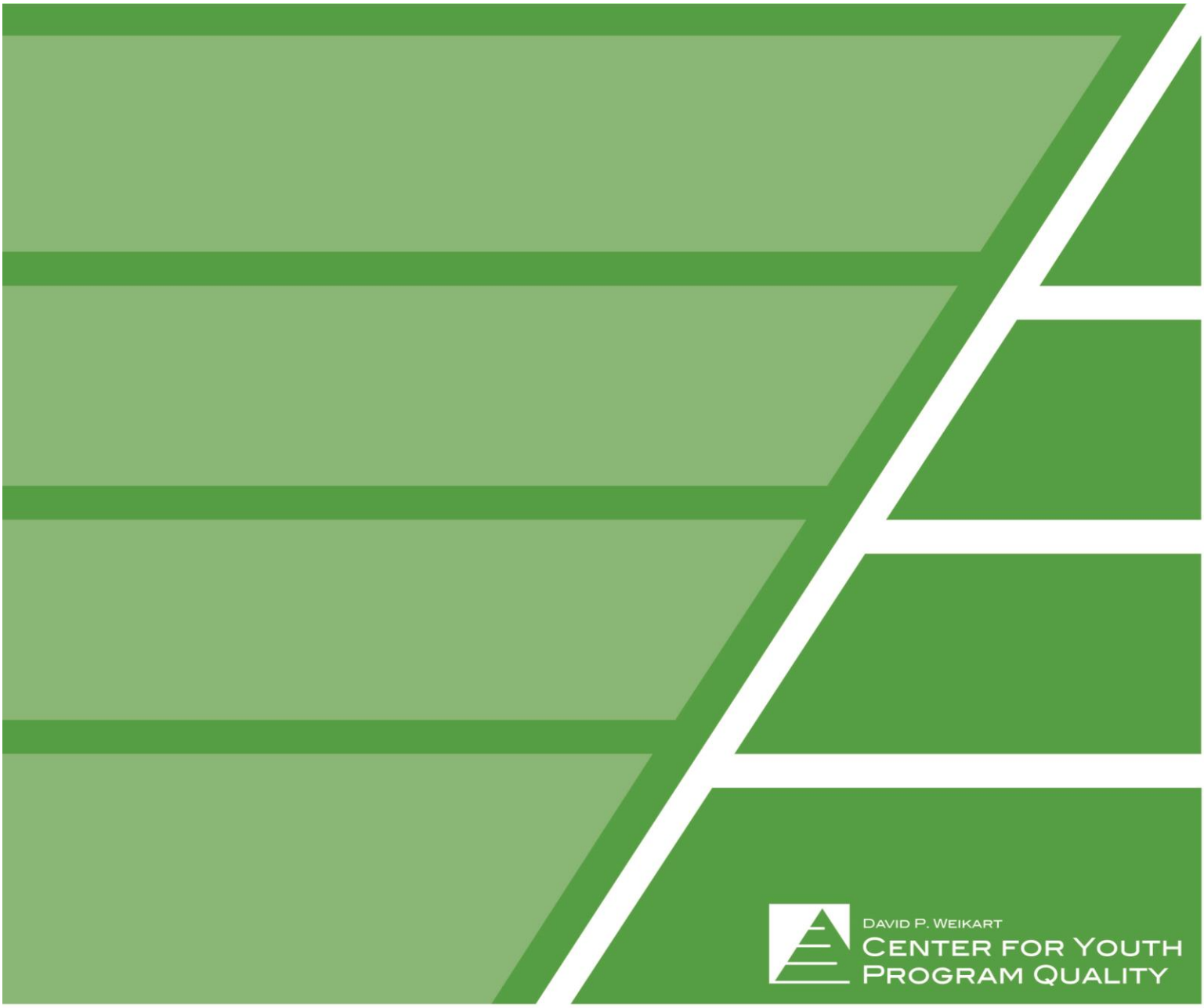
Summer Learning Program Quality Assessment

2013 Phase I Pilot Report



The David P. Weikart Center for Youth Program Quality is a division of the Forum for Youth Investment

the forum
FOR YOUTH INVESTMENT

Summer Learning Program Quality Assessment

2013 Phase I Pilot Report

February 2014

Prepared by
Ravi Ramaswamy, Anna Gersh, Samantha Sniegowski, Gina McGovern & Charles Smith
The David P. Weikart Center for Youth Program Quality, a Division of the Forum for Youth Investment

Introduction

A growing evidence base suggests that summer learning programs can benefit students, including positive effects on student achievement, reduction in summer learning loss, and bridging the opportunity gap between disadvantaged and advantaged students (Augustine, McCombs, Schwartz, & Zakaras, 2013). However, not all summer learning programs have resulted in positive outcomes for enrollees; research studies and best-practice literature show that effective programs provide opportunities for more time in settings that provide high-quality academic instruction as well as a mix of opportunities to build social and emotional skills, to experience cultural/arts enrichment, and to eat healthy food and have physical exercise. To be effective, summer programming needs to be high-quality, and students need to enroll and attend regularly (McCombs, Augustine, & Schwartz, 2011).

In 2013, the National Summer Learning Association (NSLA), the field leader in summer learning, and the David P. Weikart Center for Youth Program Quality (Weikart Center), the field leader in quality improvement system-building, collaborated to raise the quality of summer learning across the country. In this *proof-of-concept* pilot, the project team produced measures, trained raters, and piloted a performance feedback model geared at quality improvement in summer learning programs in Grand Rapids, MI; Oakland, CA; and Baltimore, MD.

The purpose of this report is to describe the concept that was tested for a summer learning quality assessment process, explain how it was implemented in this pilot, share feedback from participants and results from the data, and provide findings and recommendations for future implementation and development.

Summer Learning Program Quality Assessment Pilot

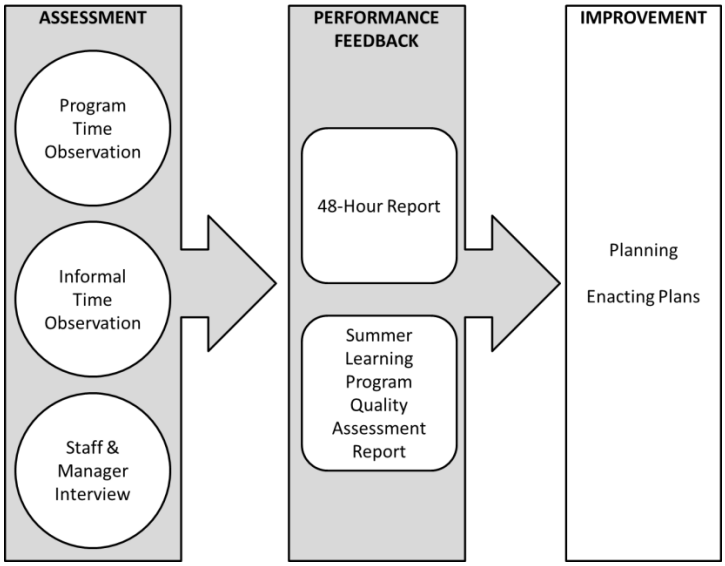
The purpose of this proof-of-concept pilot was to (1) integrate elements of existing Weikart Center and NSLA tools as a scalable quality improvement intervention for use in summer learning programs, and (2) to implement this process in a small number of programs in order to better understand the efficacy of the integrated product.

The Summer Learning Program Quality Assessment (PQA) was piloted in three cities: Baltimore, MD; Grand Rapids, MI; and Oakland, CA. A total of 16 program sites participated in the process. All of the participating sites were familiar with the Weikart Center's and the National Summer Learning Association's tools and processes. In addition to participating in the pilot process for the Summer Learning PQA, the sites in Grand Rapids also conducted a standard program self-assessment process using the Youth Program Quality Assessment (Smith & Hohmann, 2005).

The pilot consisted of the following sequence as depicted in Figure 1: collection of data through observer ratings and interviews, performance feedback in the form of a 48-hour report and quality assessment data report, and planning for and enacting program improvement.

Figure 1.

The Summer Learning Program Quality Assessment Pilot



The implementation process consisted of the following sequence of elements. For a detailed description of these elements, please see Appendix A:

1. Project Kickoff
2. Summer Learning PQA Assessor Training
3. Assessment Visit
4. 48-Hour Report
5. Summer Learning Program Quality Assessment Report
6. Planning with Data
7. Ongoing Technical Assistance

The Summer Learning PQA and the protocol for its application were developed to meet the unique expectations intrinsic to summer programming. First, the measures needed to require very little time commitment from front-line staff. Second, the assessment, reporting, and improvement

needed to be able to fit within the tight timeframe of a summer program, often less than six weeks from start to finish. Third, for this pilot, an additional goal was to design a process that provided continuity for school-year programs already using the Youth PQA as part of a continuous quality improvement process.

To meet these purposes, the Summer Learning PQA combined three pre-existing assessment measures: The Weikart Center's Youth PQA and Camp Program Quality Assessment (Akiva, 2010) and NSLA's Comprehensive Assessment of Summer Programs (CASP)[1]. One of the key innovations of the Summer Learning PQA involved adjusting the program self-assessment process, a standard element of the Youth Program Quality Intervention (YPQI)[2], to more closely match the external assessor-driven administration of the CASP tool, which takes into account the constrained timelines of many summer programs. Rather than frontline staff doing the bulk of observations, external assessors conducted site visits, observing one full program day, interviewing staff, and creating a narrative report within 48 hours. For a complete list of the Summer Learning PQA measures, see Appendix B.

Evaluation of the Pilot

Once the sites completed implementation of the pilot, feedback was collected from pilot program staff and assessors. An online survey was administered to eleven site managers, and telephone interviews were conducted with ten assessors. In addition, an analysis was conducted of the Summer Learning PQA data from 18 Summer Learning PQAs that were completed at 16 sites.

A summary of findings for the pilot evaluation are provided below. Additional analyses are presented in the appendices. Appendix C describes results from the participant survey. Appendix D describes results from the Summer Learning PQA assessor interviews. Appendix E provides an analysis of the data collected. Finally, Appendix F describes the Weikart Center's standard measurement development process.

[1] For more information see http://www.summerlearning.org/?page=program_improvement.

[2] The Youth Program Quality Intervention (YPQI) is a data-driven continuous improvement model for afterschool systems. A cluster-randomized trial of the YPQI demonstrated a cascade of positive effects beginning with the provision of standards, training, and technical assistance, flowing through managers and staff implementation of continuous improvement practices, and resulting in effects on staff instructional practices (Smith et al., 2012). For more information, and to read the full report, please visit www.cypq.org/ypqi.

Findings and Recommendations

Survey responses and interviews with assessors confirmed that the pilot was successful – the project team was able to execute the development, training, and supports needed to implement the Summer Learning PQA in the field. Reponses also clearly indicate that key elements of the process were relevant and useful to programs' day to day work.

Key findings from the pilot evaluation are:

1. Overall, sites described the Summer Learning PQA/YPQI as a positive experience.
2. Overall, administrative staff felt prepared to engage in the process.
3. One hundred percent of site managers and external assessors described the 48-hour report as useful.
4. Site managers noted they planned to use the final data report to inform improvement efforts in subsequent programming.

Key findings that indicate areas for improvement are:

1. Idiosyncrasies of particular sites point to a need for more clarity and structure for the data collection protocols given to assessors for both observation and interviews.
2. Some interview questions were ambiguous to both assessors/site managers.
3. Site managers noted that changes during summer programming were difficult given the short timeframe, but the 48-hour report was useful in bringing potential improvements to light.

Based on the findings, it is recommended that a next phase of implementation be conducted with the following goals:

1. *Refined Intervention Design*: Phase I of the pilot served as a means to test the efficacy of the Summer Learning PQA when implemented during summer programming. With positive results from the proof-of-concept pilot, the next step is to test implementation of a more complete continuous quality improvement intervention design. This effort should begin with a clear logic model that determines the key components of effective summer learning, as well as key organizational support necessary to implement the continuous quality improvement method.
2. *Revised Summer Learning Measures*: With a summer learning logic model in hand, measures should be refined and developed to align as directly as possible with the desired outputs of summer learning programs. Feasibility of implementation of the

measures in the summer learning context should be a primary consideration, as well as reliability of the data produced and validity of continuous quality improvement methods which are aligned with the measures.

3. *Increased Sample Size*: In order to better understand how the Summer Learning PQA is performing from a reliability and validity perspective, it is important to get a more robust sample of data to examine. This could include having multiple observations per site, pairing raters at a site, or expanding to a larger sample of sites.
4. *Increased Opportunities for Performance Feedback*: Improvement happens when staff members are empowered to change their behavior by learning new skills. A more effective continuous quality improvement method would increase the number of opportunities for performance feedback and improve the targeting of performance feedback to specific organizational roles. These increased opportunities could be achieved through additional training at the beginning of the summer (e.g., Instructional Coaching for managers), integration of the PQA report with the 48 hour report, a second round of assessments (e.g., after the 48-Hour report), inclusion of team-based self assessment at programs with capacity, or structured learning community calls or webinars.
5. *Assessor Training Revision*: In order to improve the preparation of quality assessors, the Summer Learning PQA training should be improved by making the support materials more specific for summer learning, increasing the amount of time trainees spend learning the Summer Learning PQA rubrics, and completing guidance for best practices when implementing the assessment protocol.

By implementing a proposed Phase II and developing a full Summer Learning Program Quality Intervention (PQI) process, both the Weikart Center and NSLA could be in a position to offer comprehensive Summer Learning PQI supports to their client base at large by 2015.

Appendix A: Description of Summer Learning PQA Implementation Elements

1. *Project Kickoff* – Following site recruitment, a webinar introduced site managers to the Summer Learning PQA measures and external assessment process. The Summer Learning PQA is the primary performance metric for the pilot.
2. *Summer Learning PQA Assessor Training* – This training was delivered to previously endorsed reliable raters for the Youth Program Quality Assessment (PQA)[3] who were recruited in each pilot site. The training provided information and skill-building on new measurement rubrics included in the Summer Learning PQA, an external assessment protocol, and how to create the 48-Hour Report.
3. *Assessment Visit* – Assessors visited each site to rate the quality of summer learning program practices through two observational forms- instruction time and program time - and one interview form focused on organizational practices and policies..
    a. Observational Ratings: Assessors viewed structured, skill-focused sessions such as academic classes, arts workshops, or sports drills (Instruction Time) as well as less structured, informal times such as meals, transitions, or greetings and closings (Program Time). For each site, assessors were instructed to view at least two Instructional Time sessions and at least 30 minutes of Program Time. The observational rating for instructional and program time focus on instructional practices and adult-child interaction.
    b. Interview Ratings: Assessors interviewed one manager-level staff person as well as one direct service-level staff person. Both staff members were asked the same set of questions about mission, professional development, and other policy-level aspects of the program. The interview data assesses systems and processes that support youth participants.

4. *48-Hour Report* – Each site received a one-page narrative report from the assessor within 48 hours of the assessment visit. The report gives the assessor an opportunity to cite several strengths that were observed and make a limited number of suggestions for small program improvements that can be enacted quickly and relatively easily by program staff.

[3] The Youth Program Quality Assessment is a validated instrument designed to measure the quality of youth programs and identify staff training needs. See (Smith and Hohmann, 2005)

5. *Summer Learning Program Quality Assessment Report* – The Weikart Center processed each site's observational and interview data into a report that disaggregated scores across the three major components of the tool: program time, instructional time, and interview.
6. *Planning with Data* – Site managers attended a Planning with Data workshop at the end of their summer session in order to learn how to interpret the PQA report and develop improvement plans that targeted both policy and point-of-service quality. These plans were incorporated into the school year programing as appropriate but were intended to be fully executed in the planning and implementation of the 2014 summer session.
7. *Ongoing Technical Assistance* – Network leads from each participating city received ongoing technical assistance support via phone and email from their Weikart Center project managers as well as NSLA staff.

Appendix B: Summer Learning PQA Measurement Structure

Table B1 provides a list of the Summer Learning PQA measures organized in descending order by Form, Domain, Scale, and Item. The table also provides the range, mean, and standard deviation of the data collected for each item at the pilot sites.

Table B1

Measure/Domain/Scale/Item

INSTRUCTION TIME (N=19)	Range	Mean	SD
Supportive Environment	**3.60-5.00**	**4.37**	**.40**
Staff provides a welcoming atmosphere.	*3.67-5.00*	*4.62*	*.54*
Staff greet youth	1.00-5.00	4.00	1.46
Staff are warm and respectful	5.00-5.00	5.00	0.00
Staff use positive body language	3.00-5.00	4.87	.50
Session flow is planned, presented and paced for youth.	*4.20-5.00*	*4.69*	*.30*
Session starts and ends on time	3.00-5.00	4.62	.80
Materials and supplies ready on time	5.00-5.00	5.00	0.00
Sufficient materials for all	5.00-5.00	5.00	0.00
Staff explain activities clearly	3.00-5.00	4.52	0.00
Appropriate time for activities	3.00-5.00	4.37	.88
Activities support active engagement.	*2.50-5.00*	*4.25*	*.72*
Youth engage with materials or ideas	3.00-5.00	4.87	.50
Youth talk about activities	1.00-5.00	3.87	1.31
Activities balance concrete and abstract concepts	2.00-5.00	4.68	.79
Activities include tangible products or performances	1.00-5.00	3.50	1.59
Skill Building: Staff supports youth in building skills.	*3.20-5.00*	*4.31*	*.65*
Learning focus is linked to activity	1.00-5.00	4.29	1.20
Staff encourages youth to try new skills	1.00-5.00	4.25	1.18
Staff models skills	1.00-5.00	4.04	1.39
Staff breaks down tasks	2.00-5.00	4.37	1.02
Support for struggling youth	3.00-5.00	4.73	.70
Staff supports youth with encouragement.	*3.00-5.00*	*4.06*	*.66*
Staff uses non-evaluative language	3.00-5.00	4.00	.96
Staff asks open-ended questions	1.00-5.00	3.56	1.45
Staff actively involved	3.00-5.00	4.62	.71
The staff uses youth-centered approaches to reframe conflict.	*3.00-5.00*	*4.33*	*.87*
Approach calmly	5.00-5.00	5.00	0.00
Seek input from youth	1.00-5.00	4.00	1.67
Relationship between actions and consequences	3.00-5.00	4.33	1.03
Staff follow-up	3.00-5.00	4.00	1.09

(Table B1 continues)

(Table B1 Continued)

Measure/Domain/Scale/Item

INSTRUCTION TIME (N=19)	Range	Mean	SD
Interaction	**2.17-4.33**	**3.25**	**.57**
Youth have opportunities to develop a sense of belonging.	*2.75-5.00*	*3.53*	*.68*
Get to know each other	1.00-5.00	3.22	1.47
Inclusive relationships	1.00-5.00	4.00	1.46
Youth identify with program offerings	3.00-5.00	3.39	.74
Staff publically acknowledge achievements	1.00-5.00	3.50	1.54
Youth have opportunity to work collaboratively.	*1.00-5.00*	*3.55*	*1.38*
Opportunities to work collaboratively	1.00-5.00	3.83	1.65
Interdependent roles	1.00-5.00	3.62	1.85
Shared goals	1.00-5.00	3.20	1.62
Youth have opportunities to develop leadership skills.	*1.33-4.33*	*2.70*	*.76*
Group process skills	2.00-5.00	3.75	1.06
Opportunities to mentor	1.00-5.00	2.50	1.41
Opportunities to lead a group	1.00-3.00	1.87	1.02
Youth have opportunities to partner with adults.	*2.00-5.00*	*3.21*	*.83*
Staff share control with youth	1.00-5.00	2.93	1.81
Staff provide an explanation	1.00-5.00	3.50	1.15
Engagement	**1.33-4.00**	**2.92**	**.81**
Youth have opportunity to make plans.	*1.00-4.00*	*2.43*	*1.15*
Opportunities to make plans	1.00-5.00	2.56	1.31
Planning strategies	1.00-5.00	2.31	1.19
Youth have opportunities to make choices based on interests.	*1.67-4.50*	*2.94*	*.85*
Content choices	1.00-5.00	3.18	1.51
Process choices	1.00-5.00	2.70	.93
Youth have opportunities to reflect.	*1.00-5.00*	*3.37*	*1.16*
Youth have opportunity to reflect on what they are doing	1.00-5.00	3.18	1.72
Youth reflect in multiple ways	1.00-5.00	2.68	1.25
Youth provide feedback	1.00-5.00	3.62	1.36
Youth present to group	1.00-5.00	3.50	1.71
Instructional Total Score	**2.65-4.42**	**3.51**	**.50**

(Table B1 continues)

(Table B1 Continued)

Measure/Domain/Scale/Item

PROGRAM TIME (N=8)	Range	Mean	SD
Safe Environment	**3.80-5.00**	**4.6**	**.32**
Psychological and emotional safety is promoted.	*2.00-5.00*	*4.5*	*.95*
Positive emotional climate	3.00-5.00	4.54	.83
Lack of bias	1.00-5.00	4.5000	1.15
Physical environment is safe and free of health hazards.	*4.67-5.00*	*4.97*	*.08*
Environment is free of health and safety hazards	5.00-5.00	5.00	.00
Clean and sanitary	5.00-5.00	5.00	.00
Ventilation and lighting	5.00-5.00	5.00	.00
Temperature	3.67-5.00	4.91	.33
Appropriate emergency procedures and supplies are present.	*2.50-5.00*	*4.13*	*.66*
Emergency procedures	1.00-5.00	4.28	1.48
Fire extinguisher	1.00-5.00	3.57	1.65
First-aid kit	1.00-5.00	3.42	1.15
Other safety equipment	5.00-5.00	5.00	.00
Supervised entrances	5.00-5.00	5.00	.00
Supervised access to outdoor space	3.00-5.00	4.75	.70
Program space and furniture accommodate activities.	*4.00-5.00*	*4.85*	*.29*
Sufficient space	3.00-5.00	4.79	.58
Suitable space	3.00-5.00	4.87	.50
Furniture	3.00-5.00	4.87	.50
Physical environment can be modified	3.00-5.00	4.87	.50
Healthy food and physical activity are provided.	*3.50-5.00*	*4.60*	*.53*
Drinking water is available	5.00-5.00	5.00	.00
Food and drinks are plentiful	5.00-5.00	5.00	.00
Available food and drink is healthy	3.00-5.00	4.75	.68
Program dedicates at least 30 minutes per 3 hours of programming daily for physical activity for all youth	1.00-5.00	3.68	1.81

(Table B1 continues)

(Table B1 Continued)

Measure/Domain/Scale/Item

PROGRAM TIME (N=8)	Range	Mean	SD
Program Climate	**3.39-5.00**	**4.12**	**.54**
Program fosters a positive, youth-centered culture.	*2.00-5.00*	*3.25*	*.96*
Program environment is characterized by an intentional, consistent behavior management style led by both staff and youth throughout the entire program day	3.00-5.00	4.12	1.02
Staff and youth hold each other accountable to a behavioral contract	1.00-5.00	3.62	1.20
Program creates a spirit of community and pride among all participants and staff	1.00-5.00	3.62	1.20
Program has principles that set culture through continuous communication of key ideals, strengths or talents	1.00-5.00	3.12	1.54
Program incorporates themes, entertainment or aspects of program culture into most of the unstructured time	1.00-5.00	3.25	1.23
Decorations in most activity spaces and common areas represent current participant work and program themes	1.00-5.00	3.12	1.54
Staff show friendliness with youth.	*3.50-5.00*	*4.80*	*.43*
Staff mainly use a warm tone of voice and respectful language	3.00-5.00	4.70	.68
Staff generally smile, use friendly gestures, and make eye contact	3.00-5.00	4.87	.50
Staff are always attentive to youth when approached	3.00-5.00	4.87	.50
Staff appear to like the youth they're working with	3.00-5.00	4.75	.68
Staff actively circulate and interact with youth	*3.00-5.00*	*4.31*	*.83*
Staff circulate (and spread out if multiple staff) to interact with every youth (in groups or individually) at some point during the program day	3.00-5.00	4.62	.80
Staff interact one-on-one at least once with every (or almost every) youth during the program day	3.00-5.00	4.25	1.00
Staff are actively involved with each youth	3.00-5.00	4.25	1.00
Each youth experiences personal attention from staff	3.00-5.00	4.12	1.02
Total Score	**3.73-4.88**	**4.37**	**.34**

(Table B1 continues)

(Table B1 Continued)

Measure/Domain/Scale/Item

MANAGER AND STAFF INTERVIEWS (N=24)	Range	Mean	SD
Purpose: Program sets annual goals for youth and for the organization that drive a continuous cycle of data collection, evaluation and quality improvement program has evidence that it is meeting its goals and the needs of stakeholders	2.30-4.00	3.34	.44
Recruitment and Enrollment	2.00-4.50	3.64	.71
Youth Outcome Goals	2.50-4.50	3.30	.59
Goal Measurement	2.00-4.00	3.30	.62
Data Collection Methods	2.50-4.00	3.30	.62
Stakeholder Feedback	2.50-4.00	3.16	.48
Planning: Program is designed to allocate enough time, staff and resources to promote positive academic and developmental youth outcomes.	2.34-3.88	3.10	.50
Continuum of Programming	1.00-4.00	2.90	.82
Total Hours of Programming	2.00-4.50	3.23	.79
Adult to Youth Ratio	1.00-4.00	2.43	.94
Proactive Planning	1.50-4.50	2.82	.975
Youth Input	1.00-4.00	2.89	1.04
Use of Certified Teacher	2.00-4.50	3.66	.67
Backward Planning	2.00-5.00	3.14	.94
Unit and Lesson Plan Framework	1.00-5.00	3.23	1.01
Staffing: Program provides extensive opportunities for staff training and development before and during the program.	1.70-4.00	3.07	.75
Staff Training	1.00-4.50	3.06	.97
Staff Collaborative Event Planning	1.00-4.50	2.96	1.09
Support for Non-certified Teachers	1.00-4.50	3.06	1.20
Staff Meetings	1.00-4.00	2.96	1.04
Staff Observation Feedback	1.50-4.00	3.26	.92
Partnerships: Program builds and maintains strong linkages with partners, including community organizations, the public school system and government agencies, that are supportive of its mission and goals.	1.33-4.00		
Partner Staff Collaboration	1.00-4.00	2.81	1.19
Year-Round Contact with Families	1.00-4.00	2.80	1.01
Relationship-building with Families	1.00-4.00	2.85	.96
Instructional Context: Program assesses young people's assets and needs early in the program and develops individualized strategies for meeting program goals.	2.13-4.00	3.20	.60
Youth Assessment	2.00-4.50	3.10	.92
Individualized, Tailored Instruction	1.00-4.50	3.14	1.00
Forward-Thinking Activities	1.00-4.00	3.36	.81
Program Principles	1.50-4.00	3.13	.71
Interview Total Score	**2.13-3.94**	**3.09**	**.55**

Appendix C: Feedback from Summer Learning PQA Participants

A thirteen-question online survey was sent via email to site managers in the three pilot networks asking what they felt about the specific elements of the Summer Learning PQA pilot process and the nature of the support provided. Eight site managers responded to three general satisfaction/projected use questions:

1. How helpful was the Summer Learning PQA Scores Report?
2. Was the 48-hour report helpful?
3. Did you execute changes as a result of the suggestions contained in the 48-hour report?

Seven of the eight respondents found both the SL-PQA Scores Report and the 48-hour report to be either "helpful" or "very helpful", selecting the most positive scale rating available for both questions. Seven of the eight respondents also answered "yes" (given yes/no options) to the third question regarding the projected use of the suggested improvements in the 48-hour report. Detailed text responses follow.

Kickoff. On a three-point satisfaction scale (Yes, Somewhat, and No), the nine site managers who responded to this question found the kickoff helpful. Of these, seven site managers responded with "yes" and two site managers identified the process as "somewhat" helpful.

48-hour report. A total of eight site managers responded to the question asking about the utility of the 48-hour report. Using the scale (Yes, Somewhat, No), seven said the report was helpful and one respondent said the report was "somewhat" helpful. Seven site managers stated that they executed changes as a result of the report. This is consistent with the external assessor responses which suggested that immediate feedback from site managers was generally positive, the only caveat being that shorter programs, like Oakland, might not be able to use the information right away. One assessor also expressed concern that programmatic changes, especially changes in staffing over time might influence the potential use of the report.

Site managers were also asked about the potential sustainability of the changes suggested in the report. The full text responses are provided below.

Table C1

Site Manager Reports on Sustainability of the Changes to the 48-hour Report

"The 48 hour report was helpful because it gave suggestions that could be implemented immediately."
"I believe the changes are very sustainable because we will be deliberate with making sure we implement the recommendations."
"It was great having the report so we can make the changes necessary."
"Very sustainable, we made an improvement plan we plan on utilizing in future program years."
"The 48 hour report was great! I thought it was very helpful! We adjusted to do the changes. Some changes will occur next year."
"It is difficult to say at this point as programming takes place at locations that we are not at each day. We do visit sites however it will be over time before we will be able to determine whether or not improvements are in line with recommendations."
"The idea here was to work on changes that were small, that could be implemented right away, or that were necessary in program improvement. They are sustainable because we see them as a progression in program quality."

Summer Learning PQA Report. All six site managers who responded stated that the Summer Learning PQA Report was helpful. When asked for recommendations for improvement, four respondents stated they had no suggestions or that the report served their needs. Full text responses are listed below:

Table C2

Site Manager Reports on Utility of the Summer Learning PQA Report

"I actually was pleased with the entire process."
"It was great. We need board members on the team."
"The language is not in my brain yet. It takes time"
"We never received our copy of the report that was scored."

Improvement planning. One site manager described the Improvement Planning process as "Excellent", while five site managers described the process as "Very Good" and one identified the process as "Good". Among the text responses, none of the respondents described the process negatively, however when asked for obstacles, respondents described the process as time consuming and difficult to implement in such a short time span. Suggestions for improving the process are displayed in Table C3 below.

Table C3

Site Manager Reports on the Improvement Planning Process

"We need time for planning for next year a lot sooner than ever!"
"[It was] Hard to implement change in a 6 week program, even with a 48 hour report."
"It was time consuming, but helpful"
"That the hard copy form used coincides with the format for inputting online."

Site managers were asked how they might carry over or transition their improvement plans into their afterschool programs.

Table C4

Site Manager Reports on How to Transition Improvement Plans into Afterschool Programming

"I will begin with reviewing the improvement plan process with staff and then incorporate it in our daily lesson plans."
"We are YPQA-ing our Library Scholars program!"
"Use improvement plan in making next year's plan."
"We are in constant eval[uation] mode and that is helpful"
"Site Coordinators will go over it with their staff and they'll work together to bring forth plans to make the improvements."
"We will refer to the data and information received in planning for next summer."

Site manager recommendations and additional thoughts. Site managers offered few additional recommendations or comments. One respondent suggested greater emphasis on the 48-hour report and one requested lesson guidelines aligned with the Youth PQA measure (complete responses are listed in Table C5 below).

Table C5

Site Manager Recommendations and Additional Thoughts

"Take the time and let the names sink into your brain before judging the program."
"Do it and ask a lot of questions"
"That it is very helpful for those new to YPQA and a review for those experienced in it but not necessary."
"Focus on the 48 hour report and we should do those sooner, maybe after the first 1 -1 1/2 of program."
"I would like a lesson plan guideline that follows the YPQA format."
"It was my first time and I'm looking forward to leading the way to become an external evaluator"

Appendix D: Assessor Interviews

The following is a selection of responses from the assessor interviews. These responses have been selected to illuminate both relevant findings and areas of potential improvement and focus for the second phase of implementation. Although assessors' overall response to the pilot was positive, they were also able to provide suggestions for improving the process. One area that garnered special attention was time management. Eight assessors described the process as "long". Of those, seven attributed this feeling to a lengthy observation period, while four assessors attributed it to a lengthy interview process. The quotes in Table D1 below are from pilot evaluation assessor interviews that characterize the Summer Learning PQA observation experience.

Table D1

Assessor Reported Experience with the Summer Learning PQA

"I think there was enough time to do the tasks. Like to do the observation to do the writing of the 48-hour report; that was not a difficult thing to complete."
"… I was observing the program for a long time. And so there definitely were times I didn't necessarily need to be there but due to requirements of the assessment there is waiting around and whatnot. That was a little bit of a challenge."
"I had to do four visits in the 18-day summer program, so it was remarkably time consuming. There was no option of requesting more time because it had to get done. It was challenging because that last week of programming students were prepping for showcase so it really didn't feel like we were getting a good snapshot of programming because in the first week you want to give programs a chance to ramp up."
"I guess overall, I felt like my observation time of instruction time was more than enough, maybe a little bit much because I sat through two, full hours of instruction and it was pretty intensive since throughout that time I took a lot of notes but most of the notes I was taking were the same."

All ten assessors described the 48-hour report as a strength of the process. The quotes below are from the pilot evaluation assessor interviews that characterize the Summer Learning PQA 48-hour Report and the interview process:

Table D2

Assessor Reported Strengths of the 48-hour Report

"One of the strengths of the process we did this summer versus the traditional YPQA is the 48-hour report or some type of report back to the site that has a little more narrative… [the sites] are obviously hungry for feedback and I didn't have any real way to provide that other than through the scores. So I think having some type of written report and it doesn't have to be long but just some type of communication was the strength of this process."
"Actually, I really like the 48-hour report. The ability to give that quick feedback is very beneficial to the site. I look forward to writing it because I understand how critical it is to get that feedback during a summer program. You know it's short and you're observing for quite a long time. I think it helps to implement that feedback right away."
"One of the sites I saw on the Monday of the last week of programming, so I really doubt they did much with the feedback….Because they were so focused on ramping up for the family showcases at the end of the week, I really doubt they did anything with the feedback. At another site, the site coordinator had found out that she was not going to be returning to the program in the fall and so I'm not sure that her [motivation to follow through] was there. For the other two sites they internalized the feedback, they agreed with me, but my sense was [they] were more focused in getting through the summer."
"I think that the site directors, the on-site staff, appreciated the opportunity to answer questions about their program and the policies and training and stuff like that. People are always a little nervous so it was nice for them even though it wasn't necessarily about programming; it was about policy and readiness and things like that but I think they still appreciated that opportunity to talk about their program as part of the process."

Overall, assessors described the process positively. When asked for recommendations, several noted that observers new to the process would benefit from additional review of the observation tool to facilitate effective use of observation time. One respondent suggested recording the sessions to have a document available for reference during reporting. Several respondents provided the specific feedback highlighted in Table D3.

Table D3

Assessor Feedback on the Summer Learning PQA Process

"I think that I really enjoyed the process and I almost wish we could do that on a regular basis and sit down with the department head and those who are running the program and really, really talk about these things that benefit really everybody from top level all the way down."
"My recommendation would be to read the tool multiple times and make a quick list of things to take notes about that you can reference… so that while I'm sitting and observing I could reference the list and then remind myself of what I need to be looking for."
"Something I did that was really helpful when I did my assessment is I actually just took my iPad and recorded the interviews I did and so I didn't have to sit there and write everything to remember it. I did go back when I went to fill out the assessment and listened to their answers. I think that was a good use of time and it helped me when I went back to do them and had actual information on what they said."
"[This process was unique] because you have a program director interview and the teacher interview. At two of the sites I got teachers because they taught as part of the summer learning program but they were teaching. The school and afterschool were very combined at two of the places so they were school teachers who were working at it. So I was talking to them and talking to the program director about the planning that they do with the PBO and the planning they do with the school and how it was aligned came out because of the [interview] questions."

Appendix E: Quantitative Analyses of Summer Learning PQA and Interview Data

Quantitative analyses were conducted on the Summer Learning PQA observation and interview data in order to get a general sense of how the items performed.

Table E1 displays the descriptive statistics for the Instruction Time and Program Time sections of the Summer Learning PQA as well as the interview items. There is a broader range of scores in the Instruction Time section of the tool than the Program Time. This is not surprising, as many of the items in Program Time were drawn upon the Safe Environment domain in the Youth PQA – in which scores tend to range between four and five.

Table E1

Descriptive Statistics for Summer Learning PQA Domains

Measure/Domain	Range	Mean	SD
Instruction Time (N= 19)			
Supportive Environment	3.60-5.00	4.37	.04
Interaction	2.17-4.33	3.25	.57
Engagement	1.33-4.00	2.92	.81
Program Time (N= 8)			
Safe Environment	3.80-5.00	4.6	.32
Program Climate	3.39-5.00	4.12	.54
Manager and Staff Interviews (N=24)			
Program Purpose	2.30-4.00	3.34	.44
Program Planning	2.34-3.88	3.10	.50
Program Staffing	1.70-4.00	3.07	.75
Program Partnerships	1.33-4.00	2.78	.90
Program Instructional Content	2.13-4.00	3.20	.60

Note. The three measures of the Summer Learning PQA are organized by domain, scale and item. The Instruction Time Measure is organized into 3 domains, 13 scales, and 44 items. The Program Time measure is organized into 2 domains, 8 scales, and 34 items. The Manager and Staff Interviews are organized into 5 scales and 25 items. For a complete list of mean scores by measure, domain, scale and item, see Appendix C.

While the data set generated from the observational ratings provides information about sites, the sample size and measurement plan do not provide a design that is highly informative regarding issues of reliability and validity.[4]

Correlation

A principle components analysis rendered factor scores for the three measures used in the Phase I intervention. The Instructional Total Score (the average combined total of the Support, Interaction and Engagement domains of the Summer Learning PQA) was included without the Conflict Resolution scale. Because actual conflict must be observed to score the Conflict Resolution scale, and because there was little score-able conflict observed during the Summer Learning PQA program assessments, the Conflict Resolution scale was omitted from this analysis. Inclusion of the scale, given the high degree of missing data, would have unfairly represented the Instructional Total Score and its correlation with the other measures.

Factor scores. Factor Scores among the three measures were not highly correlated. This is likely due to the nature of pilot efforts, which are typically burdened by inconsistent implementation and low sample size. This was the case for the Phase I pilot. The proposed Phase II implementation will present an opportunity to revisit these analyses, given a standardized implementation process informed by the challenges faced by the participating sites during Phase I. See Table E5 for a complete correlation matrix with all ten domain scores.

Table E2

Correlated PCA Factor Scores: SL-PQA, Program Time, and Manager Interviews

	ITS	Program Time	Manager Interview
ITS	1	.620	-.044
Program Time	.620	1	.057
Manager Interview	-.044	.057	1

Note: Average N= 15. No correlations were statistically significant, given p≤.05

[4] Although much is known about the reliability and validity of the Youth PQA. See for example Smith and Hohman (2005), Smith et al.(2012), Naftzger (2011). For an overview of our approach to reliability and validity of observation-based measures see Appendix F of this report.

A correlation matrix (Table E3) was created for all ten domains of the three measures. There were no significant correlations among the domains. The reasons for this may be wide range of scores and or issues related to pilot implementation.

Table E3

Correlation Matrix for 10 Domains: All Measures (Instructional Time, Program Time, Manager Interviews)

	Support	Interaction	Engagement	Safe Environment	Program Climate	Purpose	Planning	Staffing	Partnerships	Instructional Context
Sup.	1	.724	.602	.179	.207	-.207	.124	.231	-.118	.269
Int.	.724	1	.457	.051	-.046	-.238	-.154	-.028	-.504	.031
Eng.	.602	.457	1	.059	.037	-.443	-.342	.101	-.064	.052
Safe	.179	.051	.059	1	.234	-.361	-.128	.125	.125	.221
P.C.	.207	-.046	.037	.234	1	.344	.322	.328	.501	.505
Pur.	-.207	-.238	-.443	-.361	.344	1	.729	.453	.511	.542
Plan.	.124	-.154	-.342	-.128	.322	.729	1	.721	.510	.582
Staff.	.231	-.028	.101	.125	.328	.453	.721	1	.739	.793
Part.	-.118	-.504	-.064	.125	.501	.511	.510	.739	1	.749
Ins. Con.	.269	.031	.052	.221	.505	.542	.582	.793	.749	1

Note. No significant correlations were identified among these domains

Cluster Profiles

Cluster analyses (Ward Method) were conducted to establish low, medium, and high scoring sites. Prior work with similar measures has suggested that identification of scoring profiles can be of great benefit in planning future implementations. However, these analyses showed little differentiation among cluster profiles. This may suggest a wide range of instructional scores or, as in the correlation analyses, implementation and sample issues associated with the pilot effort inhibited representative cluster profiling. Among the three cluster profiles identified, none emerged as distinctly low or high. Program Time demonstrated higher scores in all profiles. Interview data demonstrated lower overall scores; however interview data was evaluated on a 4-point scale instead of a 5-point scale, as was Program Time data so Interview scores may not be fairly represented in this analysis. Normed scores were also evaluated, but the solutions produced were uninterpretable, suggesting wide instructional variation, pilot issues, or both.

Figure E1.

Cluster Profile Membership for Instruction Time, Program Time, and Interviews

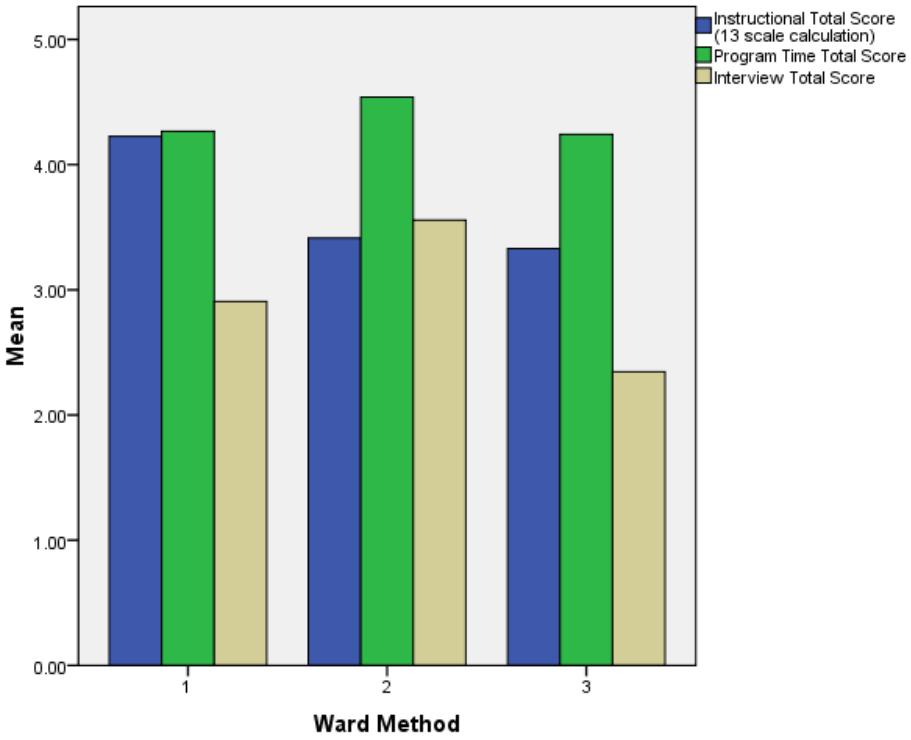


Table E4

Descriptive details of cluster profiles

Cluster Group	Number of Observations	Percent of Total
1	5	33.3
2	7	46.7
3	3	20.0
Total	15	100.0

Note. Cluster analysis was conducted using the Ward Method (Ward, 1963) The Ward Method is also known as the minimum variance method. It is intended to create homogeneous clusters with a minimum within group variance. This method is the default method in SPSS.

Appendix F –Approach to Observational Measurement Development

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, 2012) 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 following:

1. 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.
2. 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 better understood as indexes.
3. Facets of data collection – items, raters, time of day and year, programs, and interactions of these facets, may introduce substantial error into quality scores. These sources of unreliability can only be detected with data collection designs that "cross" raters.
4. 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 Youth 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 development of observational measures consists of the following steps:

1. *Content and Substantive Validity* – Which instructional practices are important and where can an observer see them? Both measures and data collection methods can be adjusted to maximize opportunities to observe instructional practices of specific interest. This step involves literature review, consultation with expert practitioners, drafting items, empirical analyses to see how items group, and asking practitioners when and where we may see these practices.

2. *Reliability* – Do multiple raters produce the same score? Our goal in this step is to maximize inter-rater reliability at the item level. Our primary analytic tools include qualitative analysis of rater reflections on the meaning of language in items, percent perfect agreement, and intraclass correlation coefficients (ICC). Internal consistency as a measure of reliability for multi-item scales is only appropriate for reflective scales. In a reflective scale, each item is theorized to "reflect" a latent construct – with interchangeability of items assumed – and any item should provide a reflection of the underlying construct; the latent construct is assumed to "cause" the item responses. For observation-based measures of behavior, however, groupings of items are most often formative in the sense that the items add up or "form" the composite score.9 Scores for formative measures are best constructed as sum scores or indexes, and are best evaluated by reference to inter-rater reliability (measures of internal consistency are not appropriately applied).
3. *Convergent Validity* — Are observation-based scores associated with other relevant measures? Convergent validation demonstrates how quality scores relate to other measures implicated by our theories of organization and child-level change (See Appendix D). Because relationships between fine-grained measures of teacher behavior (e.g., planning or reflection) are (a) not specified clearly by research and (b) likely to be context dependent10, we are frequently interested in point-in-time relationships between a total score (e.g., the setting features many good staff practices) and other policy and theory relevant constructs such as teacher education and youth engagement. Guided by theory, we employ both linear and pattern-centered analytic methods to investigate point-in-time patterns of association.
4. *Contribution of Methods to Unreliability* - How do facets of data collection method produce measurement error? Following steps 1-3, we use techniques drawn from generalizability theory to understand systematic error associated with several facets of data collection method including items, raters, time of day and year, program type, and interactions among facets. Analysis of variance methods estimate true score and error based on data collection designs that "cross" the several facets of method and use methods that maximize score reliability (e.g., more raters, more days).

Although only summarizing our approach to reliability and validity, these steps support recommendations for use of observation-based measures in lower stakes circumstances for performance feedback and continuous improvement.

References

Akiva, T. (2010). The Camp Program Quality Assessment (PQA): Development and pilot study. Ypsilanti, MI: Forum for Youth Investment.

Augustine, Catherine H., McCombs, Jennifer Sloan, Schwartz, Heather L., & Zakaras, Laura. (2013). *Getting to Work on Summer Learning: Recommended Practices for Success* RAND Summer Learning Series,

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.

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

Naftzger, N., Vinson, M., Manzeske, David, & 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.

Raudenbush, S.W., & Sampson, Robert J. (1999). Ecometrics: Toward a science of assessing ecological settings, with application to the systematic social observation of neighborhoods. *Sociological Methodology, 29*, 1-41.

Seidman, E. (2012). An emerging action science of social settings. *American Journal Community Psychology*.

Smith, Charles, Akiva, Tom, Sugar, Samantha A., Lo, Yun-Jia, Frank, K. A., Peck, Steve C., & Cortina, Kai. (2012). Continuous quality improvement in afterschool settings: Impact findings from the Youth Program Quality Intervention study. Ypsilanti, MI: Forum for Youth Investment.

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

Ward, J. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58, 236-244.