THE UNIVERSITY of NORTH CAROLINA
GREENSBORO

Tiered Quality Rating Improvement System (TQRIS)
Program Quality Measure Development Project:
Early Childhood Quality Improvement Pathway System (EQuIPS)
Final Report

Submitted to

North Carolina Department of Health and Human Services Division of Child Development and Early Education

January 30, 2017

Table of Contents

	Page Number
Introduction/Overview	**3**
Grant Personnel and Advisory Committee Members	**7**
Development Process Literature Review State Standards Item Generation	**10**
Pilot Testing Pilot 1 – Summary Report Pilot 2 – Summary Report Pilot 3 – Summary Report	**18**
Conceptual Framework and Model Overarching Principles Key Practices	**25**
Overview of the Early Childhood Quality Improvement Pathway System	**27**
EQuIPS Large-Scale Pilot	**30**
Initial Results from Large-Scale Pilot	**41**
Program Scoring Recommendations and Considerations	**51**
Training Recommendations	**53**
Website Overview and Recommendations for Use of Technology	**54**
Continuing Work (Next Steps)	**56**
References	**57**

Introduction/Overview

The purpose of the Program Quality Measure Development Project was to create a new measure of quality for use in a Tiered Quality Rating and Improvement System (TQRIS) that would allow for a broader understanding of quality in early care and education programs than is captured with most measures currently in use. Many current measurement approaches emphasize the classroom as the unit of analysis with little attention to the programmatic systems that underlie classroom performance. The promotion of programmatic strategies to ensure high quality implementation and continuous quality improvement is essential for improving the quality of early care and education programs. Implementation science provides a framework to conceptualize the foundational processes necessary to undergird high quality practice in early care and education programs. Using the program as the primary unit of analysis, the new measure embeds competency, leadership, and organization drivers as means to promote effective practice over time. The integration of a comprehensive definition of quality combined with a framework for implementation was the intent of the project.

Organizational Structure

Three states participated in this work, with North Carolina serving as the lead state for the project. North Carolina and Delaware were funded through the Round 1 Race to the Top Early Learning Challenge grant process. Both states participated in all stages of the project. A third state, Kentucky, participated in the beginning stages with support from the state Division of Child Care and University of Kentucky. The development of the measure was a highly collaborative process between the grant team members and especially between North Carolina and Delaware which were involved in all the phases of the project.

To facilitate consideration of important issues and accomplishment of key tasks associated with developing and launching a new measure, both a University Consortium and a State Administrators group were formed. The University Consortium consists of researchers from the University of North Carolina at Greensboro, the University of Delaware, and the University of Kentucky. Through contracts from their respective state agencies, members of this consortium were responsible for the technical process of developing and pilot testing the measure. This included content and organization, item choice, measurement strategies, scoring, and methods of reporting results. UNC Greensboro led the Consortium. The Consortium sought expert advice and review at multiple points during the measure development process, focused on such topics as the content of items, measurement and program assessment strategies, regulatory considerations, and implementation strategies.

State Administrators were responsible for identifying broad priorities for the project consistent with state contexts and the requirements of external funders including their ELC contracts, and for communicating these priorities to University Consortium members as subcontractors. The State Administrators group considered the broad characteristics of a measure that would be useful in their own TQRISs as well as those of other states, and contextual factors at the state and national level likely to affect the successful implementation of the measure. They also considered how ownership of the measure would be defined and how it would be made available, given the collaboration by multiple states, and the fact that the measure must be freely available according to the terms of the ELC grant. A Memorandum of Agreement was executed among participating states to further define common understandings about the goals of the project, roles and responsibilities of the parties, and decision-making processes. As the lead agency, the North Carolina Division of Child Development and Early Education facilitated communication and coordination between the University Consortium and the State Administrators. To advise them about issues, the State Administrators convened an Advisory Committee. The Advisory Committee met four times over the course of the project, either face-to-face or through phone conferencing and webinar technology. Each meeting consisted of a project update and discussion of timely topics related to the development and use of the new measure. The Advisory Committee provided feedback and posed questions that were highly beneficial in the development process of the measure.

Early Childhood Quality Improvement Pathway System (EQuIPS)

The work from this project led to the development of the initial prototype of the Early Childhood Quality Improvement Pathway System (EQuIPS), which provides a review and assessment system for use in early childhood programs. Programs' performance and assessed quality will result in a portrait based on various sources of evidence gathered through multiple methods such as observations, document review, and interviews included in EQuIPS. It is anticipated that this portrait will be used in varying ways within a TQRIS and help identify pathways for programs to improve child care experiences for children and families. The focus of EQuIPS is on children's experiences in early education and care settings, what teachers offer to facilitate those experiences and how programs are designed and organized to enhance those experiences. This measure: 1) integrates aspects of program quality not emphasized by any single measure in use by states; 2) includes structural and process features at the program and classroom level that have been linked to positive child outcomes; 3) integrates implementation science components that allow for an assessment of administrative and programmatic processes that provide the foundation for quality and continuous quality improvement; and, 4) captures a range or continuum of program quality through multiple methods. Central to the

design of the measure is the ability to maintain efficiency and reliability and avoid redundancy within TQRISs, thus the measure was intended to focus on aspects of quality not assessed by other regulatory agencies, build upon the foundation of licensing, and recognize the variety of ways in which child care programs can demonstrate quality.

The development of EQuIPS was designed from the beginning in our contract to be an iterative process that allowed for many opportunities to refine items for each component of the measure (i.e., observation, interviews, and document review). Thus, there was an intentional emphasis by the researchers to include many more items in the pilot studies than would be possible in a final version of the instrument. The intention was to test a large number of items and then, through feedback from statistical analyses, qualitative data from research team members, and expert review, to delete and refine items until the strongest and most complete quality constructs built from valid and reliable items were developed.

Goals for the Report

This report outlines the development of EQuIPS to date including the process of generating items from the literature and research based in early childhood education along with the subsequent pilot studies used to refine items and measurement methodology. The attached measure represents the initial version of EQuIPS that was tested in the large-scale pilot study, results from the large-scale pilot and recommendations for scoring EQuIPS. The report concludes with considerations and recommendations for next steps in refining items and observation methodology, training considerations, as well as the use of technology with EQuIPS. Planned next steps currently underway are also included. Table 1 shows the Outcomes from the original contract with DCDEE which were all successfully completed and references where the specific information can be located in the report as well as data sources.

Table 1. Outcomes worksheet from original contract with page references to locate related information.

	A	B	C	D
	Projected Outcomes for Proposed Activity & Target Achievement Date	**Identify the Type of Objective for Projected Outcome**	**Data Source**	**Baseline or Current Year Projected Results**
1	Collaborative team (including key staff and personnel, project partners and multi-state consortium members) for the Measure Development Project: August 2012	Performance	List of invited participants Timeline of team meetings	See page 7
2	Initial item pool for new measure: March 2013	Performance	List of items for each core domain	See page 16
3	Summary report of Pilot 1: August 2013	Performance	Data collected during Pilot 1	See page 18
4	Summary report of Pilot 2: April 2014	Performance	Data collected during Pilot 2	See page 20
5	Summary report of Pilot 3: August 2014	Performance	Data collected during Pilot 3	See page 23
6	Summary report of Large-Scale Pilot: June 2015	Performance	Data collected during the Large-scale Pilot	See page 30
7	Final report and recommendations on the new measure: December 2015	Performance	Data collected during the literature review, Large-scale Pilot, and expert review	For recommendations see complete report
8	Delivery of final Program Quality Measure December 2015	Performance	Measure	See EQuIPS measure in attached documents

Grant Personnel and Advisory Committee Members

EQuIPS was developed and tested through a grant team comprised of faculty, graduate students, QRIS experts in a multi-state collaboration between the University of North Carolina at Greensboro, the University of Delaware, and the University of Kentucky. The development of the measure was informed by input and feedback from an advisory committee made up of a variety of experts in the field of program evaluation, measurement, and early childhood education.

Grant Team

Martha Buell – University of Delaware, Co-PI

Deborah Cassidy – UNCG, Co-PI

Danielle Crosby – UNCG, Co-PI

Rena Hallam – University of Delaware, PI

Linda Hestenes – UNCG, Co-PI

Alison Hooper – DE State Anchor

Karen La Paro – UNCG, PI

Sharon Mims – UNCG, Co-PI

Rebekah Pierro – Research Assistant

Mary Lee Porterfield – Data Collection Specialist

Beth Rous – University of Kentucky, advisor

Lia Rucker – NC State Anchor

Rhonda Sawyer – Research Assistant

Christine Skrobot – DE State Anchor

Yudan Wang – Project Coordinator

Advisory Committee

Peggy Ball - Technical Assistance Specialist, ICF International & National Center on Child Care Quality Improvement

Leigh Bolick - Director, Child Care Services, South Carolina Department of Social Services

Gerry Cobb - Director of State Services, BUILD Initiative

Jana Fleming - Director, Herr Research Center, Erikson Institute

Yolanda Garcia - Director, E3 Institute, WestEd

Dan Haggard - Deputy Director, Early Childhood Services, New Mexico Children, Youth & Families Department

Ivelisse Martinez-Beck - Senior Social Science Research Analyst and Child Care Research Team Leader, Office of Planning, Research & Evaluation, Administration for Children & Families, US Department of Health & Human Services

Anne Mitchell - President, Early Childhood Policy Research

Debra Pacchiano - Director of Research to Practice, Ounce of Prevention Fund

Carla Thompson - Vice President for Program Strategy, WK Kellogg Foundation

Kathy Thornburg - Director of Center for Family & Policy Research, University of Missouri and Assistant Commissioner of Early & Extended Learning, Missouri Department of Education

Albert Wat - Senior Policy Analyst, Education Division, National Governors Association for Best Practices

Martha (Marty) Zaslow - Senior Scholar, Child Trends and Director, Office for Policy & Communications, Society for Research in Child Development

Advisory Committee Meetings

June 2013, Meeting at Professional Development Institute, Introductory Updates on Development Process

November 2013, Meeting at NAEYC, Project Update

March 2015, Via Email, Practice Profiles

June 2015, Webinar, Updates on Pilot Work and Plans for Large Scale Pilot

Expert Reviewers

Edna Collins, formerly with N.C. Department of Health and Human Services, Division of Child Development & Early Education

Beth Rous, Professor, University of Kentucky

Bobbie Weber, Faculty Research Associate, Oregon State University

Toni Porter, Principal, Early Care and Education Consulting

Juliet Bromer, Research Scientist, Herr Research Center for Children and Social Policy, Erikson Institute

Development Process

The development of EQuIPS began with an extensive review of the empirical and theoretical literature around child development and early childhood education, with particular attention focused on studies linked to child outcomes. The research team also used early learning and development standards from multiple states to guide efforts in identifying aspects of children's learning and development that were important and should be supported in high quality programs. The review of the literature led to the generation of hundreds of items that were subsequently revised and refined through a series of proposed pilot studies. Each aspect of this process will be summarized next.

Literature Review

The search for literature began by dividing the grant personnel into teams and assigning each team one developmental domain. The domains included: Language/Literacy, Social-Emotional, Physical and Health, Cognitive, Approaches to Learning, and Creative Expression. The literature search was focused on articles which included child outcomes for children ages birth through 8-years. Academic databases including ERIC, EBSCO, and PsycINFO were searched, and articles published between 2000 and 2012 were included in the search. After the database searches, the table of contents for key early childhood education journals were searched as well as reference lists from identified articles. Funded projects and major research and policy reports were also included in the literature review. Later, the literature search was expanded to ensure studies focused on inclusion, family engagement, and diversity were reviewed.

Articles were charted using a web based data system to record specific aspects of each study including characteristics such as sample size, age level, location, program characteristics, domain, data analyses, research design, major relevant findings, and initial thoughts related to items based on the article. A total of 930 unique articles across all areas were coded and charted on the website. The website interface allowed for easy review of literature by each of the key aspects coded for a study. For example, teams could go back and pull only the studies which focused on toddler social development or longitudinal studies with cognitive outcomes. This flexibility in search allowed the researchers to move back and forth between item development and the literature findings.

Summary and Accomplishments - These elements of the process provided a foundation for grounding the development of the measure in research and from a perspective of positive child outcomes. The literature review and iterative process also allowed the inclusion of the major developmental and learning areas for item construction. The initial generation of items over 100 provided a broad perspective of the range of practices and strategies used in child care.

State Standards

To ensure the inclusion of areas across multiple states, state standards from Kentucky, Delaware, and North Carolina were reviewed. The figure below (Figure 1) illustrates the plan for the literature review process and the tables show the alignment of Infant and Toddler Early Learning Standards (Table 2) across the three states and the alignment of Preschool Early Learning Standards (Table 3) across the three states. The alignment of the state standards helped the project team ensure they were focused on key developmental areas of young children's learning.

Figure 1. Plan for Literature Review and Standards Leading to Item Development

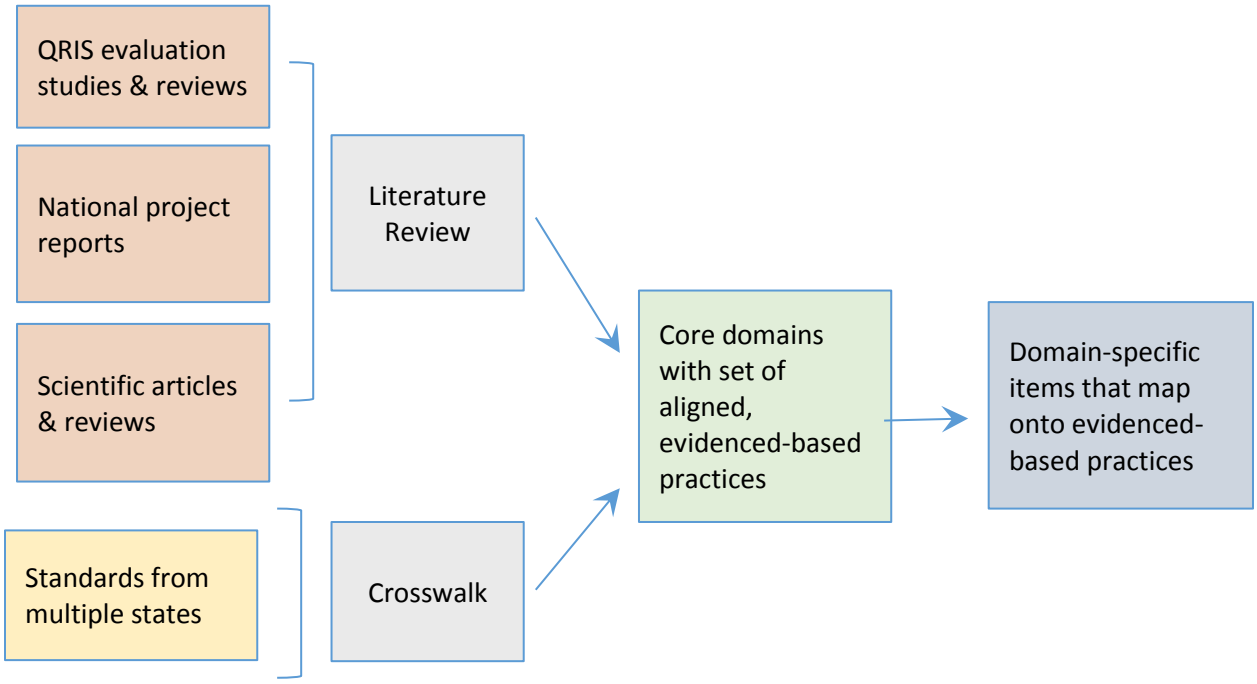

Table 2. Infant Toddler Early Learning Standards Alignment—North Carolina, Delaware, and Kentucky

North Carolina Infant Toddler Foundations	Delaware Infant Toddler Early Learning Foundations	Kentucky Birth to 3 Early Childhood Standards
Language Development and Communication	**Language and Literacy**	**Communication**
Receptive Language (Hearing/Listening/Understanding)	Receptive Language	Standard 2: Demonstrates listening and observing skills and responds to the communication of others.
Expressive Language (Talking/Communicating)	Expressive Language	Standard 1: Demonstrates communication skills in order to express self.
Early Literacy	Emergent Literacy	Standard 3: Demonstrates interest and engages in print literacy materials.
	Emergent Writing	
Emotional and Social Development	**Social Emotional**	**Social Emotional**
Developing a Sense of Self	Self-Awareness	Standard 2: Demonstrates sense of self.
Learning about Feelings	Self-Regulation	
Developing a Sense of Self with Others	Attachment/Social Relationships	Standard 1: Demonstrates trust and engages in social relationships.
Health and Physical Development	**Physical Development and Health**	**Motor**
	Sensory	
Fine Motor/Small Muscle	Fine Motor	Motor Standard 1: Demonstrates motor skills in daily activities and adaptive/self-care routines.
Gross Motor/Large Muscle	Gross Motor	
Self-Care	Health Awareness and Practice	
Physical Health and Growth		
Safety Awareness		
Cognitive Development	**Discoveries**	**Creative Expression**
Creative Expression		Standard 1: Demonstrates interest and participates in various forms of creative expression.
		Cognitive
Sensory Exploration and Discovery	Sensory Awareness	Standard 1: Explores the environment to gain information.
Concept Development and Memory	Memory	
	Cause and Effect	
	Spatial Awareness	
Problem Solving	Curiosity and Problem Solving	
Social Connections		
Approaches to Learning		
Curiosity and Eagerness	Curiosity and Problem Solving	
Confidence, Risk-Taking and Problem Solving		
Wonder and Delight		

Attention, Effort and Persistence	Attention and Persistence	
Imagination, Creativity and Invention	Play	

(Domains are shaded, sub-domains are not. North Carolina has uses "Standards" in place of sub-domains.)

Table 3. Preschool Early Learning Standards Alignment—North Carolina, Delaware, and Kentucky

North Carolina	Delaware Preschool	Kentucky 3 and 4
Language Development and Communication	**Language and Literacy**	**Language Arts**
Receptive Language	Receptive Communication	Standard 2: Demonstrates general skills and strategies of the listening and observing processes.
Expressive Language	Expressive Communication	Standard 1: Demonstrates general skills and strategies of the communication process.
Foundations for Reading	Emergent Reading	Standard 3: Demonstrates general skills and strategies of the reading process.
Foundations for Writing	Emergent Writing	Standard 4: Demonstrates competence in the beginning skills and strategies of the writing process.
Emotional and Social Development	**Social Emotional Development**	**Health and Mental Wellness**
Developing a Sense of Self	Self-Concept	Standard 1: Demonstrates health/mental wellness in individual and cooperative social environments.
	Self-Regulation	
Developing a Sense of Self with Others	Social Relationships and Cooperation	
Health and Physical Development	**Physical Development and Health**	**Physical Development**
Motor Skills	Fine Motor	Standard 1: Demonstrates basic gross and fine motor development.
	Gross Motor	
Self-Care	Health Awareness and Practice	Health and Mental Wellness Standard 1: Demonstrates health/mental wellness in individual and cooperative social environments.
Physical Health and Growth		
Safety Awareness		
Approaches to Learning	**Approaches to Learning**	
Imagination, Creativity and Invention	Initiative and Curiosity	
Aesthetic Sensibility		
Curiosity, Information-Seeking, and Eagerness		

Persistence, Attentiveness, and Responsibility	Engagement and Persistence	
Risk-Taking, Problem-Solving, and Flexibility	Reasoning and Problem-Solving	
Pondering, Processing, and Applying Experiences		
Cognitive Development	**Mathematics**	**Mathematics**
Mathematical Thinking and Expression	Number and Operations	Standard 1: Demonstrates general skills and uses concepts of mathematics.
	Geometry and Spatial Sense	
	Patterns	
	Measurement	
	Data Analysis	
Cognitive Development	**Science**	**Science**
Scientific Thinking and Invention	Sensory Awareness	Standard 1: Demonstrates scientific ways of thinking and working (with wonder and curiosity).
	Scientific Exploration	
	Scientific Inquiry	
	Scientific Knowledge—Living Things	
	Scientific Knowledge—Non-Living Things	
	Scientific Knowledge—Earth and Sky	
	Scientific Knowledge—Environment	
Cognitive Development	**My Family, My Community, My World**	**Social Studies**
Social Connections	My Family and My Community Culture	Standard 1: Demonstrates basic understanding of the world in which he/she lives.
	Past, Present and Future History	
	Places and Spaces—Geography	
	Working Together—Government and Communities	
	Money and Resources—Economics	
Cognitive Development	**Creative Expression**	**Arts and Humanities**
Creative Expression	Music	Standard 1: Participates and shows interest in a variety of visual art, dance, music, and drama experiences.
	Movement and Dance	
	Visual Arts	

	Dramatic Play	

(Domains are shaded, sub-domains are not. North Carolina uses "Standards" in place of sub-domains.)

Summary and Accomplishments - This information also guided teams as they discussed the content areas within their domains and as they examined literature to begin generating items. Finally, this process supported our continued goal to make the measure applicable for use across different states.

Item Generation

Once the literature review was completed, the project team sought to develop items for the instrument that reflected practices associated with promotion of the different areas of children's learning and development from the literature. Each team of researchers generated items for their specific developmental domain. Teams then completed a preliminary review of items for overlap across domains. Items were organized in the broad categories of child participation and engagement, peer interactions, curriculum planning, child opportunities/learning potential, affective characteristics of teachers, child assessment, physical environment and materials, family engagement, and coverage of content areas (i.e. science, math, literacy, art). For example, from the Cognitive, Language, and Literacy domains a portion of the observation was focused on the conversations teachers have with children around science and math, the words that teachers use, how previous activities are used in current activities, and how community resources are included in activities related to science and math. For Social-Emotional and Approaches to Learning, teachers' responsiveness, communication with children and adults, and children's engagement were components of the observations. Since an initial review of the items indicated that the first items were primarily targeted at the classroom level, an additional team was formed to generate program level items from the literature.

During item generation, the grant team had extensive conversations around our conceptual framework (described below) as well as key areas of importance for the care and education of young children. In order to be sure the development of the measure did not neglect topics of critical importance to the field, we held a series of targeted focus groups to gather information from professionals with expertise in: inclusion, diversity/culture, family engagement, family child care homes, and infancy. Each of these topics represent unique aspects of early childhood programming which can be difficult to incorporate well into measures of quality. The five focus groups took place in different cities around the state and usually involved 2 or 3 experts and 3 members of the research team. Specific questions around measurement challenges and goals were discussed in each group. The information gathered from the focus groups was used to inform item generation as well as ideas about the best method of data collection for specific topics (e.g. observation, teacher/director/provider interview, or document review) or unique differences that need to be considered when designing the data collection strategies (e.g., infants versus preschoolers, homes versus centers, etc.).

Summary and Accomplishments - The development phase of the measure included the literature review and item generation processes. This work resulted in the literature base from which the initial item pool was generated and a lengthy list of potential items across

developmental domains. The use of teams allowed for efficiency of gathering and reviewing literature, item generation and the opportunity for review of items by outside team members for additional perspectives. The completion of this process provided an initial draft of the instrument for use in the initial pilot study, Pilot 1. The pilot studies described next, were used to test and refine items, observation methodology as well as gathering information at the program level from directors, teachers, and providers.

Pilot Testing

Pilot 1 Item Feasibility Testing (Three phases over the period of May 2013 to December 2013): Pilot study 1 focused on feasibility testing of observation items. The items which had been generated from the literature and best practice development process were tested during this pilot. The goal of the pilot was to explore the feasibility of using the items for observations and evaluate redundancy of assessment areas (i.e., were multiple items measuring the same behavior or teaching strategy).

Phase 1 of Pilot 1 focused on generating examples for environmental and teacher support regarding items in relation to children's language/literacy/math/science/cognitive development as observed in classrooms. Grant team members began working on examples and definitions for items in the item pool in preparation for use during observations.

Phase 2 of Pilot 1 focused on the evaluation of individual indicators. During Phase 1, observations took place in 69 classrooms or homes using observation items from the initial item pool. In North Carolina observations occurred in centers and family child care homes that ranged in quality from 3 to 5 stars. The centers included privately owned, corporately owned, religious-sponsored, Head Start, Lab school, and a parent co-op program. The following questions guided the work during Pilot 1.

- Is the indicator clear? Y/N
- Was this indicator observed/present? Y/N/Unsure
- For the group observed does the indicator seem to address age appropriate needs? Y/N/Unsure
- What different sources of data could be used for this indicator? (e.g., does this have to be observed or is there another source of evidence that could be considered, anything else you observed in the environment that may have related, etc.)
- Do you think this indicator is center-specific or might it work for FCCH?
- Do indicators seem repetitive, redundant? Do we have too many or too few?
- Are there important practices related to language, literacy, cognition and math/science learning that do not seem to be addressed by these items?
- What information, if any, do you need prior to the visit to make the observation effective?

Phase 3 of Pilot 1 focused on definitions and understanding the intent of items. The questions below guided this work.

- Is the item applicable? (Y/N)
- Comment on item - such as if not applicable was this due to age group, confusion about wording, type of activity simply didn't occur, etc.? Or was it because there is a phrasing or concept that is unclear
- If applicable, is it present? (Y/N)
- Did this behavior/these behaviors/events happen rarely, sometimes, frequently, or NA because presence/absence is sufficient to consider for this item?
- Does the characteristic suggested in the item vary in quality? (Y/N) If yes, give it a score from 1 - 5 with 5 being the highest/best
- List observed examples (or what was lacking/not observed) to support your score (either Y/N, frequency, or 1-5) - also any suggestions for wording changes can be included

There was also a follow-up questionnaire given to the observers to evaluate the overall environmental support and teacher support for children's social-emotional development, which was a domain area undergoing significant revision during Pilot 1.

Summary and Accomplishments - The work from Pilot 1 narrowed the item pool; items were revised based on data gathered during this pilot; and plans for the next phase of testing were developed. At the end of each phase of Pilot 1 the research team held many discussions and domain teams worked on their group of items. By the completion of Pilot 1 the item pool was somewhat narrowed down and many items had been revised based on observer feedback related to the feasibility of the item used in the observations and for assessing quality in child care. What became clear at the end of Pilot 1 was the need for a better organizing framework for the large number of items, and a stronger link between the literature base related to high quality practices in child care and the defining characteristics of items being grouped together. The importance of the grouping emerged from assessors' feedback on being able to observe related behaviors and being able to have a structure for the assessment areas**.** The recognition of the need to focus more deeply on constructs related to program level practices also came to the forefront. It was also clear that additional item testing for feasibility of observation was needed.

Pilot 2 - Development of Initial Key Practices & Item Refinement (January 2014 - April 2015):

Pilot 2 involved documenting properties of items revised from Pilot 1 and organizing items into five initial Key Practices. The Key Practices serve as an organizing structure for major areas of importance in assessing quality with supporting constructs of what types of behaviors, teaching strategies, and practices could be included in the area. The revised items (from Pilot 1) were organized based on a framework inspired by implementation science. Implementation science provides a framework to conceptualize the foundational processes necessary to undergird high quality practice in early care and education programs.

Key Practice 1: The promotion of positive child, family and staff well-being guides the design, implementation, and evaluation of the program and defines the organizational climate.

Key Constructs
1.1 Program Philosophy/Goals
1.2 Program Operations
1.3 Staff Supervision and Professional Development
1.4 Program-Family Connections
1.5 Cultural Sensitivity and Connections to Community

Key Practice 2: Children in the program spend their time in safe and healthy indoor and outdoor environments that support play and learning.

Key Constructs
2.1 Environmental Quality (Indoors and Outdoors) – Materials and Design to support learning and to promote connections
2.2 Promotion of healthy lifestyles (routines)
2.3 Access and adaptation to the environment to meet the needs of all enrolled children

Key Practice 3: Children engage in interesting and meaningful learning opportunities across curriculum areas and classroom contexts/settings
Key Constructs
3.1 Child Engagement
3.2 Quality of Activities
3.3 Quality of Routines and Transitions
3.4 Planning for learning across curriculum areas
3.5 Learning opportunities provided across curriculum areas

Key Practice 4: Children experience positive and stimulating interactions with teachers and their peers. These interactions are viewed and implemented as a central process for children's learning.

Key Constructs
4.1 Teacher-child interactions – quality, quantity, consistency
4.2 Peer-peer interactions
4.3 Teacher-teacher

Key Practice 5: Children's individualized and unique needs are met in ways that promote their healthy growth, development, and learning. Families are central in meeting these needs.

Key Constructs
5.1 Child Assessment and use of data to inform curriculum
5.2 Collaboration with other professionals
5.3 Family collaboration

<u>Video Evaluation during Pilot 2</u>
Items under Key Practices 2, 3, and 4 were evaluated via video clips of classroom activities based on revisions. The video clips represented various activities (e.g., transition, meal time, outdoor, group time) in child care classrooms of different age groups. Specifically, the following questions were addressed for each item in order to facilitate further item refinement:

1. Was this item observable in the video? (1=Yes; 0=No)
2. Could you score this item Low Medium High? (1=Yes; 0=No)
3. Is the item redundant within the construct? (1=Yes; 0=No)
4. Is the item redundant within the key practice? (1=Yes; 0=No)
5. Is this item redundant across other constructs or key practices? (1=Yes; 0=No)
6. How meaningful is this item to the construct? Rate: 1 – 5
   [1 = not meaningful, 3 = mostly meaningful, 5 = highly meaningful]
7. How meaningful is this item to the key practice? Rate: 1 -5
   [1 = not meaningful, 3 = mostly meaningful, 5 = highly meaningful]
8. Would the content of this item be better captured in interview? (1=Yes; 0=No)
9. Would the content of this item be better captured in document review? (1=Yes; 0=No)
10. Does the item need to be reworded? (1=Yes; 0=No) [comment box]

The initial item pool included strategies for assessing assessment and collaboration constructs highlighted in Key Practice 5, but these areas proved difficult to assess through observation and were moved to interview items for teachers and administrators.

Administrative Practices – Key Practice 1

Key Practice 1 was addressed in a parallel activity during Pilot 2. Guidelines were developed to evaluate program administration practices. As stated in Key Practice 1, the emphasis on program administration in addition to classroom practices made this tool a program level measure and a tool that can assist in both quality rating and quality improvement. Examples of documents were gathered from programs and rubrics were developed for evaluating documents.

Summary and Accomplishments - The initial Key Practices were developed based on principles from implementation science and grounded in the literature that had been reviewed. Based on this work and consideration of data collection and scoring, a second draft of the measure (observation + interview items) was developed. The development of these initial Key Practices and Key Constructs allowed for items to be organized and grouped into meaningful categories. The project team also connected relevant literature to the key constructs in order to eventually facilitate a pathway between the constructs being measured and support for program improvement.

Program level areas were also conceptualized within the key constructs and methods for gathering information at the program level as well as in areas difficult to assess through observation were moved to the interview.

Pilot 3 Initial Test of Full Measure & Training Materials (May 2015 – December 2015):
Pilot 3 involved developing training materials and protocols for the assessment, and testing all components (document review/interview and observation) of the instrument in programs in preparation for the large-scale pilot. Decisions were made regarding which two observers would become the 'Anchors' or master observers for NC and DE. The four designated state anchors served as the trainers and reliability checkers for the observation portion of the instrument. Twenty-six time sampling cycles of observations were conducted in North Carolina and 28 cycles were completed in Delaware by grant team members. The next phase of the pilot involved the anchors in each state completing additional paired observations using the most updated version of the observation form to establish initial interrater reliability and further refine the scoring manual and assessor training materials. This process occurred for 10 cycles in NC and 7 in DE. The anchors from each state met and completed paired observations during 13 additional cycles to verify cross-state agreement and anchor reliability.

Teacher/Director/Provider Interview - To provide information related to a program's organizational climate, family and community partnerships, and supports for teaching and learning, interviews with the program administrator and classroom teachers were conducted. Creation of these interviews was iterative and involved multiple phases of testing potential questions and various drafts of interview scoresheets. Additionally, a review of documents related to program policies, procedures and operations were included, which were developed concurrently with the interviews.

The center document review/interview rubric was developed to incorporate and evaluate concepts in the Key Practices related to program operations. These concepts foster a positive organizational climate and highlight the role of administrators in supporting best practices in classrooms and encouraging professional development as well as promoting teacher efficacy. The rubric items were grouped into four areas: 1) Staff Hiring, Supervision and Support, 2) Stability and Continuity, 3) Support for Teaching and Learning, and 4) Family Engagement and Community Partnerships. Within each area were specific items related to the construct, e.g., Methods for communicating with families, Orientation of newly hired staff, etc.

In addition, the rubric provided four levels of specific practices or procedures ranging from 1) no evidence of a practice, to 2) evidence of an emerging practice, to 3) evidence of a system in place for a practice, up to 4) evidence of multiple methods of implementing a practice. Written policies or procedures were viewed as necessary to achieve "evidence of a system". The written policies provide opportunities for greater continuity over time. Without written guidelines practices would be more susceptible to erode with hiring of new administrators

and/or teachers.

Summary and Accomplishments - In Pilot 3, final decisions were made regarding the format and procedures for each aspect of the measure and items were refined again in preparation for the large-scale pilot (pilot 4). Procedure manuals were drafted and finalized for the observations and interviews. A document review checklist was refined and tested with a small number of centers and homes. The key discussions and decisions resulting from Pilot 3 testing included:

- Testing of different intervals for time-sampling and other observation items
- Consideration of observing 1 vs. 2 teachers in a classroom – since it was hard to track 2 teachers separately, decision was to simply observe for behaviors/practices regardless of which teacher is enacting them
- Testing a picture checklist to assess indoor and outdoor environments; this was later disregarded as a strategy
- Adding teacher measures on decision-making, efficacy, and climate to the pilot
- Recommending that the global items be broken apart by frequency and quality
- Recommending that the interview focus on areas not included in the observation such as intentionality, individualization, approaches to guidance, culture, and family and community connections.
- Continuing discussion regarding data management
- Continuing discussion regarding potential scoring
- Continuing discussions on website development and inputting of EQuIPS data in addition to comparison measures being used in pilot 4; refining the web-based calendaring system to schedule observations and interviews.

Following the completion of Pilot 3, there was additional item refinement and initial consensus building to establish initial interrater reliability. Simultaneously with the pilot work, grant team members were refining and describing the intent of the measure, the organization and number of Key Practices, and solidifying the conceptual model which is discussed next.

Conceptual Framework and Model

Overarching Principles

Throughout the development of this initial draft of EQuIPS, the project team and those supporting the project have engaged in numerous discussions around the underlying conceptual framework and model supporting this important work. The conceptual model is now represented in three components with the child's experiences as the central focus, then the factors that affect the child's experiences, and finally the program administration and organizational climate as the overarching component (see Table 4). Most proximal to quality experienced by children is their actual in the moment experiences. Three guiding questions are the focus of this component: 1) where do I spend my time? 2) How do I spend my time?, and 3) How does this relate to the rest of my life? The second component includes factors that affect the children's experiences. Children's experiences are impacted by the environmental context of their center or home, classroom and outdoor space, as well as their community setting. Children's relationships and interactions with their caregivers and peers also impact the quality of their daily experiences. Additionally, children's experiences are affected by the interactions and relationships between the caregivers and between caregivers and the families in their environment. Teaching practices also directly influence children's experiences and include the strategies used to engage the group of children in the classroom or home, as well as individualized strategies for specific children. Finally, as the third component, program administration and organizational climate essentially 'surround' the center or home and provide structure in the form of policies and procedures that guide the functioning and well-being of the program.

Table 4. Conceptual Framework

Child Experiences as the Central Focus
Where I spend my time
How I spend my time
How this relates to the rest of my life

Factors that Affect Child Experiences
Environmental Settings
Relationships & Continuity
Teaching Practices
Program Administration & Organizational Climate

Overview of the Early Childhood Quality Improvement Pathway System

Based upon the conceptual framework and model for EQuIPS, the measure is designed to be a review system of program performance that utilizes observation, interviews, and document review. The system is not designed to lead to a single numeric score but rather to provide a program portrait that highlights areas of program strength and provides a pathway that will allow for professionals to support improvement and quality enhancement to ultimately benefit the children and families (see Figure 2). The intent of EQuIPS is summarized below.

Intent of EQuIPS

A review **system** of early childhood program performance and quality that provides a **program portrait** based on information from observations, document review, and interviews.

Our goal is that this portrait will be used in varying ways within a QRIS and help identify **pathways** for programs to **improve experiences** for children and families.

Figure 2. Conceptual Model Providing Overview of Measure and Final Key Practices

The information gathered through EQuIPS derives from three different sources: observation, interviews, and document review. As seen in Figure 3 the document review materials are gathered from the program and submitted for review. Center directors and teachers, or home

providers are interviewed either on site or via telephone for 30 to 45 minutes. Information gleaned from the program documents and responses to questions from interviews with directors, teachers, and/or home providers are used to inform rating on administrative rubrics. The third source of information comes from classroom or home observations. The observation procedures include both time sampling (observe for 90 seconds/code for 2 minutes) and global ratings (from 1 to 5). Observers also record the presence of any 'red flag' or concerning practices which take place during the observation. Observations take place over 'cycles' which last for 50 minutes each. Within these 50 minutes, observers complete 10 intervals of time sampling and also spend time observing and taking notes in order to score the global ratings. In the large scale pilot (information following) each classroom was observed across three cycles. The observations can take place either indoors or outdoors and have been used in infant, toddler, and preschool age classrooms, as well as in mixed age home settings. At the end of the observation cycles, the observer completes an indoor and outdoor checklist and rates the final curriculum global items (from 1 to 5). Tables 5 to 8 provide an overview of each source of evidence and primary topics addressed within each evidence.

Figure 3. EQuIPS Multiple Sources of Evidence

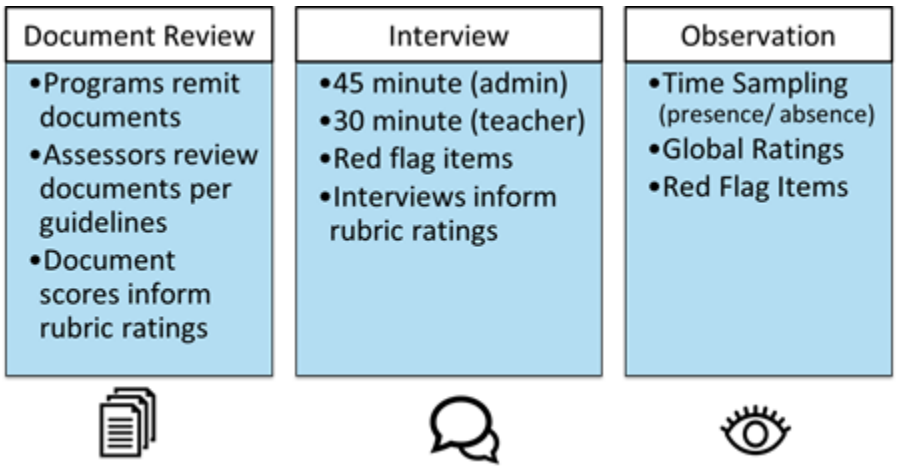

Table 5. EQuIPS Document Review Overview

EQuIPS Document Review
Policies & Procedures Related to: ➢ Staff hiring, Supervision, & Support ➢ Communication ➢ Support for Teaching &Learning ➢ Family Engagement & Community Partnerships ➢ Other Operational Policies

Table 6. EQuIPS Administrator Overview

EQuIPS Administrator Interview

Topics:

- Quality Improvement Priorities
- Staff Hiring, Supervision, Support
- Communication
- Supports for Teaching & Learning
- Partnerships with Families and Communities

Table 7. EQuIPS Teacher Interview Overview

EQuIPS Teacher Interview

Topics:

- Planning and Curriculum
- Assessment & Individualizing
- Classroom Community
- Communication with Families
- Professional Development

Table 8. EQuIPS Observation Overview

EQuIPS Observation

Time Sampling/Global Quality Indicators/Red Flags:

- Quality of Interactions
- Quality of Conversations
- Supporting Persistence and Learning in Activities and Routines
- Peer Interactions
- Curriculum & Pedagogy

Physical Environment (indoors and outdoors):

- Availability and function of space
- Presence and use of materials

EQuIPS Large Scale Pilot (January to September 2016)

The purpose of the Large Scale pilot study (pilot 4) was to better understand the functional use of the new measure in a range of program types, star levels, and locations, as well as to complete initial reliability and validity checks on the new tool. We were also interested in exploring initial scoring options for this multi-faceted measure. Assessors from both North Carolina and Delaware collected data using the prototype version of EQuIPS developed at the end of Pilot 3. North Carolina's study included both centers and family child care homes while Delaware's study included only family child care homes. The two studies will be outlined below with delineations between when procedures or measures differed across the two states. The study completed in North Carolina was done in conjunction with researchers from Frank Porter Graham (FPG) Child Development Institute at the University of North Carolina at Chapel Hill. Researchers from FPG were completing the TQRIS Validation study throughout the state during the same time period and therefore coordinating data collection across participating centers and homes was a beneficial collaboration.

North Carolina Recruitment and Participants

Participants were recruited by FPG to participate in both the NC-TQRIS Validation study and EQuIPS large scale pilot study. Programs were stratified by 1) Urbanicity (rural and urban); 2) Region (East, West, and Central) and Star rating. One hundred-seventy-five Child Care Centers and seventy-five Family Child Care Homes (FCCH) were randomly selected and recruited in North Carolina to participate in the Validation Study and a subset of this sample were recruited for the pilot of the EQuIPS measure. This combined recruiting effort led to 151 programs initially agreeing to be in the EQuIPS large scale pilot, including 99 Child Care Centers and 52 FCCH. Research personnel from FPG consented personnel at participating centers. After receiving signed consent forms, an administrative assistant from the NC project followed a recruitment script and called each center and home to set up the observation visits and discuss the procedures for the document collection and interview components of the study. Participants each received gift card incentives for their participation. Throughout the course of the 9 months of data collection 20 centers and 23 homes withdrew. The attrition of programs was a result of centers or homes closing before data collection was complete, programs deciding to discontinue participation due to the time commitment of the study, changes in administrative staff, and lack of bilingual assessors. The final NC sample included 79 centers and 29 homes stratified by location, star level, and urbanicity (See Table 9). During recruitment, each center was invited to include up to 2 classrooms each, of infants, toddlers, and preschool age children; therefore programs could have from 1 to 6 classrooms included in the study, depending on the size of the center. Across the 79 child care centers, observational

data was collected in infant (n= 43), toddler (n= 121), and preschool age (n=97) classrooms. One hundred and ninety-nine teachers and seventy-one directors were interviewed. Twenty-nine family child care home providers participated in the interviews. Across both family child care homes and centers, 63 programs provided documents for review.

Table 9. Demographic Data for North Carolina (n=108) Center and FCCH Participants

Program Characteristics	**North Carolina**	
QRIS Participation	**Centers (n=79)**	**FCCH (n=29)**
Star Level 1 & 2 (for Centers)	14%	10%
Star Rating 2 (for FCCH)		3%
Star Rating 3	16%	24%
Star Rating 4	32%	28%
Star Rating 5	38%	35%
Urbanicity of Program		
Urban	78%	76%
Rural	22%	24%
Region		
East	17%	34%
Central	54%	38%
West	29%	28%

Delaware Recruitment and Participants

Family child care providers were sampled from the population of the 705 licensed family child care homes in Delaware. The population of programs was stratified by both QRIS star level and geography to ensure a wide range of quality levels and statewide representation. The goal was to sample an equal number of providers in each of these four quality groups: Not participating in QRIS, at star level 1 or 2 in the QRIS, at star level 3, and at star level 4 or 5. Programs that were not selected for the sample were ranked within each stratum, and additional programs were added to the sample if providers declined participation, were not eligible because they were not caring for children or no longer licensed, or could not be reached. Participants were recruited through telephone by University of Delaware research staff. The research assistant used a recruitment script to assist with the call. The script included information about the purpose of the study, the type and length of program visits that would take place if the provider decided to participate, and the incentive available. A total of 391 providers were contacted to participate in the project. Of those, 121 were not interested, 177 could not be reached, and 43

were not eligible. The final sample consisted of 50 providers in the following quality groups: 9 not participating in QRIS, 17 at star level 1 or 2, 7 at star level 3, and 17 at star level 4 or 5. Table 10 shows demographic characteristics of the participating providers.

Table 10. Demographic Data for Delaware (n=50) FCCH Participants

Program Characteristics	Delaware
QRIS Participation	
Starting with Stars	14%
Star Level 2	20%
Star Level 3	16%
Star Level 4	22%
Star Level 5	10%
Not Participating	18%
Urbanicity of Program*	
Urban	
Metropolitan	92%
Micropolitan	2%
Rural	6%

* Rural-Urban status is defined by RUCA codes (Metropolitan = 1-3, Micropolitan = 4-6, Small town/Rural = 7-10)

Measures and Training

EQuIPS Observations. The observation process for EQuIPS was comprised of time sampling items, global score ratings, and a review of space and materials. Three cycles were completed in each classroom/home. Each cycle was comprised of 15 minutes of global observation, and 35 minutes of time sampling (2 minutes observing and 1.5 minutes of scoring for 10 intervals). Timing for each cycle was tracked by a timer, pre-recorded with the appropriate observation intervals that alerted the observers when to observe and when to record. In North Carolina, two cycles were observed before nap time and one cycle observed in the afternoon after nap time. In Delaware the EQuIPS observers in the homes completed three observation cycles during one morning. Space and materials were observed throughout the three cycles and scored at the end of the last cycle. Teachers or providers were asked any clarifying questions needed to score the indoor/outdoor checklist after the last cycle. To complete a cycle a minimum of two children needed to be awake.

There are 43 time sampling items in this initial version of EQuIPS. Within these 43 items, 22 items focus on how the teacher/staff interact with the children and 21 items focus on evidence

of curriculum and pedagogy. For each cycle there are 20 red flag items of concerning events that may occur during the observation cycle. The first 15 minutes in the classroom for each cycle is reserved for global observation. There are 16 global items that are scored at the end of each 50-minute cycle and 11 curriculum/summary globals that are completed after 3 cycles. There are 21 space and material items collecting information about both the inside and outside environment. (The EQuIPS score sheet and manual are included in separate documents with the report.)

***EQuIPS Observation Reliability Training*.** During Pilot 3, two "anchor" or master observers from NC and two anchors from DE completed a series of reliability checks and debriefings to ensure the observations and training process would be comparable in NC and DE. These reliability visits occurred in a preschool room, toddler room, and infant room in an early childhood center and they completed three reliability visits in family child care homes. The four state anchors were over 85% reliable on all components of the observation before assessor training began. Assessor training occurred with four NC assessors and two DE assessors. This consisted of a full day orientation on the observation form sections and observation procedures with several sample scoring activities including the use of classroom/FCCH videos, photographs, and a draft version of the EQuIPS observation form training manual. Assessors were asked to review the observation form and manual prior to the training, as well as watch a practice video and practice scoring. Correct scores were discussed as part of the orientation. The full-day orientation was followed by six days of observation based training and debriefing with an anchor in center-based child care and family child care settings. For each practice observation, pairs of or individual trainees completed the observation with an anchor, assigned scores, and then discussed the score for all items. At a minimum, assessors received training during 11-17 observation cycles in center-based classrooms representing a balance of infant, toddler, and preschool classrooms and varying quality levels and 4-5 observation cycles in FCCH settings. As part of the training process, assessors watched and coded three additional video segments. Their scores were compared against the "master code" to further establish initial interrater agreement. Interrater reliability was calculated for each observer pair based on exact agreement for the time sampling sections, red flag section, and indoor/outdoor space items. Agreement within one point on a five-point scale was used for the two sections addressing global ratings.

The average interrater reliability achieved during **training** for NC assessors was (*n*=80 cycles):

- Time sampling (average across all intervals): 87%
- Time sampling curriculum (average across all intervals): 92%
- Red flag: 98%
- Indoor/outdoor space: 87%

- Global: 88%
- Global curriculum: 90%

The average interrater reliability achieved during **training** for DE assessors was (*n*=40 cycles, 13 programs):

- Time sampling (average across all intervals): 89%
- Time sampling curriculum (average across all intervals): 94%
- Red flag: 99%
- Indoor/outdoor space: 84%
- Global: 94%
- Global curriculum: 96%

During the study anchors completed paired observations with each assessor to maintain interrater agreement. The reliability checks occurred for each assessor and anchor as needed to equal 20% of the programs included in the sample. A comparable number of reliability checks occurred for each assessor and the observation dates were distributed across the data collection time period for the EQuIPS observations (January-July 2016). For each assessor, reliability checks occurred in different age groups for center-based care and also at FCCH sites.

The average interrater reliability achieved during the **checks** for NC assessors was:

- Time sampling (average across all intervals): 90%
- Time sampling curriculum (average across all intervals): 95%
- Red flag: 98%
- Indoor/outdoor space: 89%
- Global: 96%
- Global curriculum: 97%

The average interrater reliability achieved during the **checks** for DE assessors was (*n* = 30 cycles, 10 FCCH programs):

- Time sampling (average across all intervals): 89%
- Time sampling curriculum (average across all intervals): 96%
- Red flag: 96%
- Indoor/outdoor space: 89%
- Global: 92%
- Global Curriculum: 95%

EQuIPS Interviews. The interview process for EQuIPS included an interview with the primary teacher in each classroom/home observed and an interview with the administrator. The

interviews were audio recorded, notes were taken, and typically lasted from 30 to 60 minutes. Interviews were completed either in person or over the telephone depending on teacher preference and interviewer location. The interviews targeted four key implementation processes: bi-directional communication, formal system participation, data for decision-making in the classroom, and individualization. These implementation processes were primarily embedded within the questions asked of administrators and teachers. The administrator interview had 32 questions related to 5 areas of practice in the program: 1) Staff Hiring, Supervision, and Retention, 2) Continuity and Stability for Teachers and Children, 3) Teaching and Learning, 4) Family Engagement and Community Partnerships, and 5) General Quality Improvement. The teacher interview was comprised of 23 questions focusing on areas related to both the teacher's classroom and the program as a whole. The areas related to the classroom were about: 1) Teaching and Learning and 2) Family Engagement. The program level questions in the teacher interview were about practices related to: 1) Continuity and Stability for Children in the Classroom and 2) Staff Hiring, Supervision, and Support. The FCCH interview had 38 questions around 4 areas of practice in the program: 1) General Quality Improvement, 2) Continuity and Stability, 3) Teaching and Learning, and 4) Family Engagement and Community Partnerships. Scripted clarification questions were consistently asked if the interviewee did not understand the question or if more description was needed. (The EQuIPS interview protocols are included in separate documents with the report.)

***EQuIPS Interview Training*.** Interviewers were trained using the EQuIPS Interview Procedures by grant team members who were instrumental in the development of the interview questions and rubric. The process for training interviewers included guidelines on audio-recording interviews, use of the website and submitting responses, as well as how to access resources if help was needed. Prior to conducting interviews, mock interview sessions with the trainers were completed and discussions held to resolve any scoring differences.

***EQuIPS Document Review*.** A checklist with identified areas of interest related to policies and procedures was provided to center administrators and home providers at the beginning of the project. Participants were encouraged to identify locations of specific policies within their documents for each area. Documents were picked up onsite, submitted electronically or mailed and then scanned. Submitted documents were then reviewed for the presence or absence of written policies, procedures, guidelines, etc. The areas reviewed were aligned with the items in an administrative rubric. Protocols and scoring guidelines were developed for the document review process and used as part of the training for reviewers. (The EQuIPS document review checklist is included in separate documents with the report.)

EQuIPS Document Review Training. Training for document reviews was provided by grant

team members involved in the rubric and document review checklist development. A review of written protocols that included guidelines for scoring, definitions of terms, use of the website for data entry, and resources for help was conducted. Additionally, a review of initial completed checklists was conducted with on-going email question/answer availability provided by the trainer.

Comparison Measures. In order to examine aspects of the validity of the new EQuIPS measure, data was also collected using additional measures of quality. All comparison measure data collection was completed by different assessors than those who completed EQuIPS. Observations of global quality and teacher-child interactions were conducted using the Classroom Assessment Scoring System (CLASS) and the Environment Rating Scales (ERS). In centers with more than one classroom for each age level (e.g. two infant rooms), the comparison measures were only completed in one room per age level. At the program level a shortened version of the Preschool Administration Scale (PAS) or the Business Administration Scale (BAS) was completed.

Classroom Assessment Scoring System (CLASS; Pianta, LaParo, & Hamre, 2008) provides an assessment of the quality for teacher-child interactions. Depending on the age of the children in the classroom, the Infant, Toddler or Preschool version of the CLASS was used. The preschool version of this measure has 10 dimensions which are organized into 3 domains. The Emotional Support domain includes positive climate, negative climate, teacher sensitivity, and regard for student perspectives. The Classroom Organization domain includes behavior management, productivity, and instructional learning formats. The Instructional Support domain includes concept development, quality of feedback, and language modeling. Each dimension is rated from 1 to 7 with higher scores indicating higher quality. Data collectors observed classrooms for 4 rounds of observation for 20 minutes each, followed by 10 minutes of scoring. Studies have found a link between CLASS domains and other measures of quality, such as the Environment Rating Scales (ERS) (Early et al., 2006) and child cognitive and socio-emotional outcomes (Burchinal et al., 2008; Mashburn et al., 2008). A separate version of the measure is available for classrooms serving toddlers (LaParo, Hamre, & Pianta, 2012) and infants (Hamre, La Paro, Pianta, & LoCasale, 2014). Although there is no comparable version for family child care homes, the CLASS developers suggest that the tool can be used to measure the quality of interactions in home-based settings, and it has been used in other studies (Vitello, 2014). The grant team developed a protocol to select whether the Toddler CLASS or Pre-K CLASS would be used in each program. The observer used the Toddler CLASS if the majority of children present were toddlers (15-36 months). The observer used the Pre-K CLASS if the majority of children present were preschoolers (3-5 years). If equal numbers of toddlers and preschoolers were present or if there was at least one infant present, the Toddler CLASS was

used. None of the participants in homes had only infants or only school-aged children enrolled.

Environment Rating Scales (ERS) were used to assess the global quality of child care programs. The Environment Rating Scales include three different versions appropriate for early childhood programs – the Early Childhood Environment Rating Scale-Revised (ECERS-R; Harms, Clifford, & Cryer, 1998) for center-based preschool programs, the Family Child Care Environment Rating Scale - Revised (FCCERS-R; Harms & Clifford, 2007) for family child care programs, and the Infant/Toddler Environment Rating Scale - Revised (ITERS-R; Harms, Cryer, & Clifford, 2003) for center-based programs serving infants and toddlers. These measures assess programs' structure, provisions for learning, and teaching and interactions. For this study, all the subscales on the ERS were used with the exception of Parents and Staff/Providers. Scores on the ERS can range from 1 to 7 with higher scores indicating higher quality. Studies have shown a relationship between the ERS and other indicators of program quality (Early et al., 2008) cognitive and socio-emotional outcomes (Burchinal et al., 2000, 2008).

The Program Administration Scale (PAS; Talan & Bloom, 2011) was used to measure center-based program quality across a variety of constructs including child assessment, family partnership, center operation, and staff qualification. The PAS contains 25 items grouped into 10 subscales that measure leadership, management, and administrative practices of center based early childhood programs. The instrument has a multi-use design and can be used for program self-improvement, technical assistance and monitoring, pre-service and in-service training, research and evaluation, and public awareness (Talan & Bloom, 2011). In this study, we used the following subscales: (2) Supervision and Performance Appraisal, (4) Compensation, (8) Risk Management, (11) Assessment in Support of Learning, (14) Program Evaluation, (16) Family Communications, and (17) Family Support and Involvement.

The Business Administration Scale (BAS; Talan & Bloom, 2009) was used to measure the overall quality of business and professional practices in family child care settings. Measured across a variety of constructs, the BAS contains 10 items that are measured on a 7-pt scale. When used in conjunction with FCCERS-R, it provides a comprehensive picture of quality within the learning environment and business practices that support the program. The instrument has a multi-use design and can be used for program self-improvement, technical assistance and monitoring, pre-service and in-service training, research and evaluation, and public awareness (Talan & Bloom, 2011). In this study, we used the following items: (6) Risk Management, (7) Provider-Parent Communication, (8) Community Resources, and (10) Provider as Employer.

Comparison Measure Reliability Training. To ensure high quality reliable data, data collectors were trained as stipulated by protocols established by observation measures developers. For

instance, all CLASS observers in NC and DE were certified as reliable when they demonstrated a proficiency of 85 percent or higher (within one point) interrater reliability.

All assessors using the ERS had previous training and had been deemed reliable in the ECERS-R, ITERS-R, and FCCERS-R for their state QRIS and other research projects. All assessors were required to remain at an 85% reliability level (within 1 point) and underwent extensive training on the scales before starting data collection. Assessors from DE completed two reliability visits with a state anchor from NC to ensure consistency in scoring the FCCERS-R across states. Reliability was calculated at 90% for these two visits. The PAS and BAS measures were administered by trained assessors from FPG in NC. These assessors had also met the scale requirements for training before beginning data collection. To minimize drift across all measures in the study, interrater reliability visits were conducted with 20% of classroom observations. The reliability standard was met at each reliability check for each instrument.

Participant Demographic Forms & Classroom Summary. Teachers, administrators, and providers each completed demographic forms to gather individual information, including race/ethnicity, education, experience, and length of time in the program. A classroom survey was completed by each teacher or provider to record the age levels, race/ethnicity, developmental status, and gender of children in the participating classrooms or homes.

Data Collection Procedures in North Carolina

In North Carolina, data collection occurred across multiple days depending on the size of the center or home. Since consent forms had previously been obtained from researchers recruiting the sample from FPG, the first call to the participants from UNCG was to answer questions regarding the study and set up the schedule for the observation and collection of demographic forms and administrative documents. The scheduler asked a series of questions to confirm the number of classrooms participating in the study, the ages of the children in each room, and discussed a plan for data collection. Care was taken to ensure only one data collector was in a classroom at a time unless a reliability check was taking place. The EQuIPS observers were able to collect data across two classrooms during one day. They spent 50 minutes in one classroom to complete a cycle and then switched to a second classroom for the next 50 minutes. This rotation occurred throughout the morning and afternoon until three cycles were collected across the two classrooms and the indoor/outdoor checklists and global ratings were completed. Comparison measures were completed by different trained assessors following the protocol for each measure. All comparison measures only occurred in one classroom per age level, whereas the EQuIPS observations were completed in all participating classrooms (up to 6 per center). The ERS assessments occurred over the course of a morning and the teacher interview was completed at a time convenient for the teacher. The CLASS observers completed

4 cycles of observations using the prescribed protocol. The PAS and BAS assessments were completed by trained researchers from FPG on a different day than UNCG's data collection.

The scheduler from UNCG also set up the time for each teacher and director interview. Interviews were either completed on site or over the telephone depending on participant preference and interviewer location. Each interview lasted between 30 and 60 minutes. Delaware data collectors also completed the interviews for family child care participants in North Carolina to ensure consistency across all family child care interviews. Those interviews were completed by telephone after the participants had completed all other components of the pilot. For document review, center administrators and home providers were given the option to mail, email, or have an assessor pick up copies of program documents. Each participant received a list of possible documents that could be provided (e.g. staff handbook, parent handbook, written schedule, family communications, conference forms, professional development plans, sample assessment/portfolio, etc.), and were asked to complete a checklist of policies and practices that were included within their documents. All participants received a gift card at the end of their center or home's data collection process as an incentive for participation.

Scoring of administrative rubrics for centers was completed by one EQuIPS Co-PI and one project data collector. Responses from interviews and document reviews were identified and aligned for each rubric item and progress level. Scoring was completed using online reports generated by the website showing individual responses from each data source (administrator interview, teacher interview and document review) for each item. Protocols and guidelines for determining levels of progress on rubric items were developed and included examples with specific scenarios across the multiple data sources.

Data Collection Procedures in Delaware

In Delaware, data collection occurred across three home visits: one visit for obtaining written consent and collecting program documents and two visits to complete program observations, including the EQuIPS and comparison measures. After a provider verbally agreed to participate in the pilot, the scheduler attempted to schedule the three data collection visits. Providers received a letter through mail and email listing the dates of the observation and how to prepare for the consent/document review visit. Providers received a gift card at the end of the third visit as an incentive.

The consent and document review visit was scheduled for nap time or the evening, based on the provider's preference, and occurred before the observations. First, the researcher gave the participant a consent form and reviewed the consent form with the provider. The provider

signed one copy, which the researcher filed, and kept another copy for future reference. The provider completed the provider demographic form and classroom roster form. The researcher aggregated information from the classroom roster to complete the classroom summary demographic form to ensure confidentiality. The researcher provided parent letters about the observations with instructions to send home to families before the observations occurred.

Providers then were asked if they had examples of any program documents to share with the researcher, including the following:

- Family handbook
- Program handbook (if different than your family handbook)
- Family contract
- Daily sheets
- Lesson plans
- Conference forms
- Documentation of family communications or family events
- Evidence of any partnerships/arrangements with other agencies (e.g. public schools, libraries, etc.)
- Professional development plans
- Completed screenings or child assessments

These documents were scanned or photographed, or the provider gave the researcher a paper copy to keep. Documents were uploaded to a secure server and de-identified when the researcher returned to the office.

Approximately one to two weeks after the consent/document review visit, two additional visits consisting of program observations were completed. Each visit lasted approximately four hours, including about three hours of observation and one hour of provider interview. The two observation visits occurred within one week of one another. During the observation visits, data collectors completed the EQuIPS, CLASS, and FCCERS-R observations.

Data Entry

As part of the grant project, a website was developed for scheduling purposes, data entry and cleaning, and data output. (The website design and functionality is described in more detail in a later section of the report.) Researchers from the project received training on the website and then entered all data into the site according to procedures developed for each measure. All the EQuIPS observation data, ERS scoring, and CLASS data was double checked by a second team member and any errors were corrected. Prior to all data analyses all variables were checked for missing values, means, and ranges.

Initial Results from the EQuIPS Large Scale Pilot

As described above, the multi-method, large-scale EQuIPS pilot generated large amounts of data about multiple aspects of early childhood program quality, across family child care and center-based settings. Analyses proceeded in several phases in order to evaluate the reliability and validity of individual items, their organization into constructs, the interrelationships among constructs (within and across methods) and their associations with established comparison measures. We summarize the results of these initial analyses below and discuss their implications for next steps in refining and evaluating the measure.

Item-Level Analysis of EQuIPS Observation Data

The first phase of data analysis for the observation component of draft measure involved examining the means, standard deviations, and distribution of individual time sampling, global rating, and red flag items. We found that most items demonstrated a reasonable amount of variation and were normally distributed.

Descriptive statistics were used to help identify low incidence items, as well as those with low variability. The research team discussed such items as candidates to be dropped from the measure, however, a select few were retained because of their conceptual importance. For example, teachers were rarely observed defining terms or concepts for children, yet this practice can support children's language and cognitive development and should be encouraged. Moreover, the EQuIPS measure was intentionally designed to include items with a range of "difficulty" (i.e., easier to achieve items that give programs credit for doing some of the basics, mid-range items, and more difficult to achieve items associated with higher levels of quality).

During this phase of the analysis, the research team also used information about interrater reliability and field notes from assessors to identify potentially problematic items that should be revised or deleted from the measure. For example, one of the teacher language time-sampling items captured the frequency of teachers giving directives or instructions. However, this item is conceptually different than the others (and did not correlate statistically) because high frequency would not typically be reflective of high quality. Another example is a time-sampling item capturing how often teachers provided specific and authentic feedback during the observed intervals. According to assessors, this item was difficult to score because it requires significant inference and is contingent on conversations or interactions that may be happening over an extended period of time. In the case of this item, we concluded that time-sampling is not the most appropriate method for capturing this practice, and an alternative method should be considered in the next revision of the measure.

Construct Mapping and Creation of EQuIPS Observation Subscales

During the second analysis phase of the observation data, items were organized conceptually into six subscales (related to our Key Practices 5-7), and a series of steps were taken to provide an initial examination of subscale reliability and validity. Red flag items were not included in this phase as they conceptually serve a different purpose; these are discussed further below in the scoring section.

First, the internal consistency of each scale was evaluated using Cronbach's alpha; the standardized alpha coefficient was used to account for the fact that items varied in terms of measurement scale (i.e., frequency count of 1-30 for time sampling items, and global ratings of 1-5). Item-total correlations were used to identify individual items that should be excluded to improve internal consistency. In addition, the internal reliability of each subscale was compared across cycle (1-3), child age group (infant, toddler, and preschool), and facility type (centers and homes); these analyses revealed little variation, suggesting the subscales were not more or less reliable for one group versus another. The subscales and their properties are described briefly below and summarized in Table 11.

Quality of Teacher-Child Interactions. This subscale consists of nine items (time-sampling and global ratings) that assess the frequency and quality of teachers' engagement, responsiveness, and positive support of children during interactions. Sample items include "Does a teacher address children's cues or bids for attention in a positive, accepting manner?" (time-sampling), "Teacher verbalizations are relevant to children's activities and/or interests" (global rating), and "Teacher interactions with children are warm and positive" (global rating). This subscale demonstrated an acceptable level of internal reliability ($\alpha = .88$) and was included in the next phase of analysis examining the intercorrelations among subscales and their association with the comparison measures of classroom quality.

***Enriched Language*.** This subscale consists of six items (time-sampling and global ratings) that assess the frequency and quality of teachers' support for children's language development. Sample items include a time sample of how often "a bi-directional teacher-child conversation (or interaction if the child has no/little language)" is observed, and a rating of the extent to which "teachers use rich and varied language." The internal reliability of this subscale ($\alpha = .72$) was deemed too low to reliably rank classrooms and it was not included in further analyses. A review of item and scale properties, and trainer/assessor feedback led to the decision to recommend several new items be developed to strengthen this subscale in the next version of the measure.

***Curriculum*.** This subscale consists of sixteen items (time-sampling and global ratings) that

capture the frequency and quality of children's opportunities to build knowledge and skills across a range of developmental domains and academic areas (including math, literacy, science, pretend play, music, art and physical development) through a variety of engaging activities and varied instructional strategies. The internal reliability of this subscale was high ($\alpha = .91$) and it was included in the correlational analyses.

Learning about Self and Others. This subscale includes ten items (time-sampling, global ratings, and indoor environment checklist) intended to assess the extent to which children are supported in learning how to identify, express and manage their emotions, recognize the perspective of others, engage in positive peer interactions, and whether the environment promotes community and a sense of belonging. The internal reliability of this subscale was low ($\alpha = .58$) and could not be improved by dropping any particular item. Given these results, the research team developed recommendations for revisions to item wording and new global rating items focused on how much children are supported in learning about emotions and behavior, how teachers acknowledge and incorporate various forms of diversity, and whether the environment supports meaningful connections to family and community.

***Daily Organization*.** This subscale consists of five global rating items assessing to what extent children appear to experience an environment that is predictable, child-focused and engaging, with smooth transitions between activities, and clear behavior expectations. This subscale was found to have a high level of internal reliability ($\alpha = .91$) and was included in the correlational analyses.

Environmental Foundations. This subscale contains sixteen checklist items based on observations of the indoor and outdoor environments, and whether they afforded children opportunities for easy movement and access to activities and materials, and exposure to natural elements. Given low internal reliability ($\alpha = .67$), this subscale was not included in the correlational analysis; instead the pilot results were used to generate recommendations for item revision to better capture this construct.

Table 11. Properties of EQuIPS Observation Subscales (Large Scale Pilot version)

Subscale	Item Type	# of Items	Internal Consistency
Quality of Teacher-Child Interactions	time-sampling, global ratings	9	$\alpha = .88$
Enriched Language	time-sampling,	6	$\alpha = .72$

	global ratings		
Curriculum	time-sampling, global ratings	16	$\alpha = .91$
Learning about Self and Other	time-sampling, global ratings indoor checklist	10	$\alpha = .58$
Daily Organization	global ratings	5	$\alpha = .91$
Environmental Foundations	indoor checklist, outdoor checklist	16	$\alpha = .67$

The decision was made to move forward in subsequent analyses with only the subscales that demonstrated acceptable internal reliability using the criteria of an alpha coefficient above .80 (Teacher-Child Interactions, Curriculum, Daily Organization). The items and constructs represented in the remaining subscales (Enriched Language, Learning about Self and Other, and Environmental Foundations) are still considered to be conceptually important components of observation scale; however, the large-scale pilot findings suggest that revisions are needed to improve the reliability and validity of these subscales.

The subscale inter-correlations shown in Table 12 are disattenuated to account for measurement error. They indicate strong positive relationships among all of the subscales, suggesting a high level of cohesiveness, with each contributing unique information as well.

Table 12. Correlations among EQuIPS observation subscales (N = 224)

	T-C Interactions	Daily Organization	Curriculum
T-C Interactions	1	.804	.719
Daily Organization	.804	1	.852
Curriculum	.719	.852	1

Associations between EQuIPS Observations and Comparison Measures of Quality

The third phase of analyzing the large-scale pilot observation data involved computing bivariate correlations between the three internally reliable EQuIPS subscales described above and

several established measures of early childhood care and education quality. These comparison measures are described in detail in the Large-Scale Pilot methods section.

Given that factor analyses across multiple studies in the literature have not supported the original Environment Rating Scale subscale scores, we calculate correlations based on the total ERS scores, as well as factor-based scores identified in published work (using exploratory and confirmatory factor analytic methods). Two factor-based scores were used for the ECERS-R: Language/Interactions and Materials/Activities (Cassidy, Hestenes, Hedge, Hestenes, & Mims, 2005); three factor-based scores were used for the ITERS-R: Language/Interactions, Materials/Activities, and Safety/Organization (Hestenes, Cassidy, Hegde, & Hansen, 2007); and, three factor-based scores were used for the FCCERS-R: Language/Interactions, Materials/Activities, and Organization (Schaack, Vi-Nhuan, & Claude, 2013).

The correlational analysis results are presented in Tables 13-16, and summarized here:

- Low to moderate positive correlations were observed between the EQuIPS subscales and the three CLASS domains for settings serving preschool-age children. The strongest association exists between EQuIPS: Teacher-Child Interactions and CLASS: Emotional Support, which is to be expected given the focus of each on capturing warm, positive and responsive interactions. Interestingly, the weakest association was found for EQuIPS: Curriculum and CLASS: Instructional Support. This may reflect the fact that the EQuIPS items for this subscale focus on the variety and quality of children's learning opportunities across domains and academic areas, while the CLASS Instructional Support domain focuses on what teachers are doing with respect to concept development, quality of feedback, and language modeling (constructs more closely aligned with other EQuIPS observation items not included in these correlations, such as those in the Enriched Language subscale).

- We find moderate positive correlations among all of the EQuIPS observation subscales and the ECERS-R scores for preschool classrooms. Notably, a relatively strong positive association exists between the EQuIPS Teacher-Child Interactions subscale and the ECERS-R Language/Interactions score, which is consistent with our expectations.

- In classrooms and FCC settings serving infants, we find low to moderate correlations between the EQuIPS subscales and the ITERS-R total score and factor-based scores. The lowest correlations (.1 to .2 range) are observed for EQuIPS teacher-child interactions and the ITERS-R Materials/Activities and Safety/Organization scores. That these more structurally-focused aspects of the environment are not highly related to observed interactions is not unexpected. At the same time, the Daily Organization and

Curriculum subscales of EQuIPS are moderately correlated with the ITERS-R Language/Interactions scores, which likely reflects the fact that the EQuIPS items on these scales focus on teachers' facilitation of activities, learning, and transitions.

- Finally, we observe moderate to strong positive correlations between the EQuIPS subscales and the FCCERS-R scores. Interestingly, in contrast to the findings for infant settings, the FCCERS-R score most strongly associated with EQuIPS Teacher-Child Interactions is the Organization score.

Table 13. Correlations between EQuIPS observation subscales and Pre-K CLASS domain scores

	EQuIPS: T-C Interactions	EQuIPS: Daily Organization	EQuIPS: Curriculum
CLASS: Emotional Support	.549	.440	.264
CLASS: Instructional Support	.307	.212	.142
CLASS: Classroom Organization	.384	.421	.216

Notes. The sample size for this analysis is 65 preschool classrooms and FCC programs serving predominantly preschool-age children.

Table 14. Correlations between EQuIPS observation subscales and ECERS-R factor-based scores

	EQuIPS: T-C Interactions	EQuIPS: Daily Organization	EQuIPS: Curriculum
ECERS-R Total	.470	.324	.338
ECERS-R: Language/Interactions	.623	.494	.481
ECERS-R: Materials/Activities	.445	.300	.373

Notes. The sample size for this analysis is 66 preschool classrooms and FCC programs serving predominantly preschool-age children.

Table 15. Correlations between EQuIPS observation subscales and ITERS-R factor-based scores

	EQuIPS: T-C Interactions	EQuIPS: Daily Organization	EQuIPS: Curriculum
	n=49	n=29	n=33
ITERS-R Total	.243	.439	.460

ITERS-R: Language/Interactions	.412	.563	.549
ITERS-R: Materials/Activities	.191	.167	.414
ITERS-R: Safety/Organization	.238	.505	.457

Notes. Sample sizes for this analysis vary across subscales because of pairwise deletion.

Table 16. Correlations between EQuIPS observation subscales and FCCERS-R factor-based scores

	EQuIPS: T-C Interactions	EQuIPS: Daily Organization	EQuIPS: Curriculum
	n=29	n=22	n=28
FCCERS-R Total	.437	.413	.433
FCCERS-R: Language/Interactions	.560	.307	.341
FCCERS-R: Materials/Activities	.379	.435	.507
FCCERS-R: Organization	.753	.578	.618

Notes. Sample sizes for this analysis vary across subscales because of pairwise deletion.

EQuIPS Interview and Document Review Rubric Results

The primary analyses conducted with the interview and document-based data collected as part of the large-scale pilot involved a multi-stage process of developing rubrics and scoring programs according to these criteria.

A total of 57 child care centers and 79 family child care homes (FCCH) were evaluated using rubrics developed for each specific setting. For centers, three sources of evidence were used to determine the progress made on each rubric level: 1) documents for review, 2) administrator interview and 3) at least one teacher interview. For centers, a large majority of items (80.64%) on the rubric were evaluated to be in the "Emerging" category, reflecting a presence of some beginning level of understanding or policy related to the construct but no systematic implementation. The percentage of items on the rubric where centers had written policies and reflected a systematic implementation of the construct was 8.87% with an additional 4.43% of items reflecting not only policies but multiple methods of implementing the construct. Additionally, 6.06% of the items evaluated were determined to not yet be present in the center.

For centers, the rubric items most frequently indicated to meet the higher levels were those related to Family Involvement and Methods for Communicating with Families. Additionally, the item related to use of a Variety of Teaching Strategies was frequently rated as meeting higher levels. It is worth noting that this item was more strongly influenced by interview than by

document review as compared to other items. The items noted most often as not yet being present in the center were those related to child assessment, both the staff support and the systematic storage and use of data from child assessments.

Results from use of the rubric for FCCH settings, had similarities to centers in the results. The majority of items for FCCHs also were in the beginning or "Getting Started" category (54.50%) with 25.64% of items rated as not yet present in the program. The two highest levels on the rubric reflected 5.48% of the items at the highest level with 14.38% at the second highest level. While there were some differences between FCCHs in Delaware and North Carolina, both states reflected similar patterns. For those FCCHs with items reflecting they were not here yet, the percentages were 24.77% for Delaware sites and 27.15% for those in North Carolina. For those with items indicating they were "Getting Started" at a beginning level, the percentages were 51.08% (Delaware) and 60.48% (NC) and for those items rated as meeting the highest level the percentages were 7.38% (Delaware) and 2. 15% (NC). The items most often noted as not yet being present were those related to the use of an identified curriculum and implementing developmental screenings for children.

Table 17. Distribution of rubric scores for center-based programs

Centers	Percentage of total Rubric items
Not yet present	6.06%
Emerging/getting started	80.64%
Systematic implementation including written policies	8.87%
Multiple ways of implementing	4.43%

Table 18. Distribution of rubric scores for Family Child Care Homes (FCCHs)

FCCHs	Percentage of total Rubric items	NC FCCHs	Delaware FCCHs
Not yet present	25.64%	27.15%	24.77%

Emerging/getting started	54.5%	60.48%	51.08%
Written plan/systematic	14.38%	10.22%	16.77%
Implementing	5.48%	2.15%	7.38%

Summary of Large-Scale Pilot Results

The large-scale pilot of EQuIPS provided an initial opportunity to assess multiple aspects of early childhood program quality. As noted, the pilot generated a large amount of data from multiple sources across center and home-based settings serving infants, toddlers, and preschoolers.

One of the primary goals for the analyses presented above was to assess the psychometric soundness of the classroom/home observation component, and the extent to which the current set of items represent a reliable and valid measure of classroom-level key practices that contribute to (and reflect) program-level quality. The pilot results suggest that the current version of the observation tool yields three reliable subscales--Quality of Teacher-Child Interactions, Daily Organization, and Curriculum--that are moderately correlated with established measures of classroom quality, which have some overlapping focus, but at the same time are less centered around capturing children's experiences than is intended with EQuIPS. The pilot results also suggest that there is additional work to be done to refine items that capture enriched language, children's learning about self and others, and aspects of the indoor and outdoor environments. Some of the revisions highlighted by the pilot data have to do with determining the appropriate measurement strategy for assessing areas of practices. For example, the pilot results suggest that some areas originally captured using time-sampling methods may be more appropriately measured using global ratings.

Another goal of the pilot data analysis was to develop a strategy for integrating and scoring the teacher interview, administrator interview and document review data. The current versions of the center and FCCH rubrics applied to the large-scale pilot data revealed that most programs evidenced a beginning level of implementation related to the constructs, with some evidence of understanding or practice, but without clear systems or policies in place to support full implementation. These initial findings will be informative for the next phase of refining the measure and integrating the program- and classroom-level assessments to provide a multi-

dimensional portrait of quality.

Program Scoring Recommendations and Considerations

The goal of EQUIPS is to provide a comprehensive portrait of how a program functions and performs with respect to promoting positive child outcomes by integrating measurement obtained from observation, interview, and document review. In order to produce scores or evaluations that are meaningful and useful for quality improvement as well as suitable for predicting child outcomes, the scoring system will be component based and the scores will be able to generate categorical levels of quality.

Component-Based Scores

The core components will be listed in the general rubric into which constructs in the observation, interview, and document review measures will be fitted. Within the rubric, constructs and items in the three forms of measurement will be aligned conceptually, although they may have different layouts or formats. Scores of each component will be created and then linked together dynamically to present a pathway for improvement for providers, and component scores will eventually be used to predict child outcomes. In addition, classroom-level together with program–level pathways may be produced, with the recognition that averaging classroom scores to create program scores is likely to reduce reliability by ignoring classroom–level (within-program) variability. Further, percentile information may be presented for incentive purposes when more data are collected and more analyses are conducted.

Since the components will be comprised of constructs from different sources (observation, interview, and document review), and the constructs themselves may consist of a collection of items of various formats and types, we are aware of the need to efficiently integrate the data in a statistically and operationally valid manner. Specifically, individual items will be compiled into subscale scores, subscale scores will be compiled into construct scores, and construct scores will be compiled into component scores. This is not an easy task due to the comprehensive nature of the measure. Currently, there are binary (Yes/No) items, Likert-type (1 – 5, 1 – 3) items, and checklist (whether a certain number of conditions are satisfied). Also, there are items that capture micro-level interactions together with items that capture macro-level performance. Moreover, we might give extra weight to certain items based on their conceptual importance. Given that mixing item type is likely to reduce reliability, careful consideration will be taken in the process of creating component scores. It is a balancing act among measurement reliability, theoretical necessity, and practicality of data collection.

Generating and Using the Portrait

It is anticipated that component scores will be presented graphically – as a portrait literally – in

the form of a bar chart (each component being one ‘bar’) or a pie chart (each component being one ‘slice’). As such, specific strengths and weaknesses of a program will be self-explanatory to readers of the evaluation report. Further, cut points can be created for component scores, and labels can be attached to each cut point (e.g., >20% developing, >40% acceptable, >60% performing, >80% exemplar). Such cut points may be determined conceptually at first, and may be generated statistically with sufficient data. Eventually, incentives and technical assistance can be allocated accordingly based on cut points of specific components.

Another use of the component scores is to group programs into ‘clusters’ based on the patterns of component scores. For example, certain programs tend to have high scores on all components, while others tend to have low scores on all components, and yet another group of programs tend to have high scores on a subset of components and low scores on the rest. In this way, programs with similar portraits will be identified, and such information may also be used for incentives and technical assistance.

Red Flag Items

The “red flag” items are collected and scored separately from the items indicating good quality. Similarly, various levels of “warning” can be created for summary scores of “red flag” items. These "warning" notes can be added to the portrait graphic in a similar manner as a "note" on a credit report that is presented together with the credit score. While such a "note" is separate from the calculation of the credit score, it certainly will affect creditor decisions.

Summary and Accomplishments - The large scale pilot study provided very helpful data on all aspects of this initial version of EQuIPS to assist in planned revisions. Data analyses will continue and deeper investigations of core components and constructs to support the refinement of scoring protocols and graphical representations of a program portrait.

Training Recommendations

While training materials will need to be revised, based on changes made to the measure following the most recent pilot, we believe that the basic training process sufficiently established understanding of correct interpretation and overall assessor interrater agreement. We recommend continuing to use a mixed-mode training process that includes:

- Pre-orientation reading assignments to introduce the observation form and additional written information.
- An orientation that includes an overview of each section of the observation form, assessment methods, and data collection procedures. The orientation should also include various scoring activities such as the use of video examples.
- Additional video segments with written scoring justifications to provide additional scoring practice and/or reliability training for specific sections of the observation form.
- Paired observations and debriefing with a trainer/anchor. Training sites should represent the range of program types and age ranges included in a study or QRIS.
- Detailed written information such as a manual that provides definitions and interpretation guidance for each item. Additionally, to further inform ongoing training development and clarification for assessor interpretation questions, a question and answer database with responses from the anchors should be maintained data collection.

Website Overview and Recommendations for Use of Technology

The Measure Development Project created and maintains a website (www.rttt.org) that serves as the starting point for individuals interested in learning about the project and for authorized users to enter assessment requests, upload documents, enter scores, write and generate data reports, and manage assessment schedules using a customized online calendar system.

The website was designed over the course of the project for multi-state collaboration with built-in features that grant and/or restrict access to data and website features based on member properties (e.g., audience type, certification type, university affiliation, state). Based on these properties, users can view and download many types of resources from the website (e.g., audio recordings, documents, videos, reports). Resources are grouped into meaningful categories (e.g., comparison measure documents, EQUIPS measures and forms, early learning standards, procedures, key practices) and properties for each resource are collected (such as, audience type, start/end posting dates, resource location) to allow the website to grant and/or restrict access based on member properties.

A custom calendar system allows authorized users to enter and view daily events, activities, and notes. The calendar is integrated with the assessment scheduling process so scheduled assessments appear on calendars immediately as they are scheduled. Users may click on calendar events to view information about the activity, such as the event type, event description, participants, start and end times, billable miles, travel time, and time spent writing, proofing, and training.

A custom data report system allows users to view reports that are important to their work, such as lists of pending assessments for one or more assessors, lists of people certified on a specific measurement tool, and lists of files uploaded to facilities. The report system also allows authorized users to enter scores and other information associated with most of the measurement tools used by MDP (e.g., EQUIPS: Observation form, Rubric, Teacher interview, ECERS-R, CLASS). Authorized users may also access and download data via a custom data download report that returns the requested data and data codebook describing key data characteristics (e.g., data type, range, value-labels).

With regard to data management and security, the data are housed in a Microsoft SQL database, which is behind an additional firewall layer within UNCG. Backup copies of the virtual machine are stored offsite per UNCG Enterprise-level data management policies (see, https://its.uncg.edu/services/service/enterprise-data-storage-backup- recovery-services). The data are only accessible to authorized MDP personnel through UNCG. Transmittal of data

between the web server and individual browsers is handled using secure HTTPS protocols. The data in the database are not encrypted; however, backup copies of the SQL database are encrypted and password protected. MDP uses two methods to backup database and website files. First, a database backup is created nightly using Microsoft SQL server 2008 R2 backup utilities. Cobian software is used to compress, encrypt, and password protect these backup files, as well as website source files, and then Cobian transfers the files to another equally secure server behind a firewall at UNCG. This backup method allows the database administrators to easily restore individual files and/or data tables on an as- needed basis. The types of data collected and used include information about program characteristics (e.g., permit type, size), classroom characteristics (e.g., number of children, room type, attendance), assessment information (e.g., EQuIPS scale scores, reports), teacher characteristics (e.g., gender, education, experience), and non-identifiable information about children (e.g., total number of boys/girls, ethnic origin groups, types of disabilities). No individually identifiable or confidential information about children is collected.

Recommendations for use of technology for EQuIPS include electronic data entry for observation and interview data and a document upload feature for programs as well as scoring and report generation. The data structure used by the website was designed to be scalable and flexible for long-term use. Most data tables follow a hierarchical and relational structure. This relational structure is based on the simple idea that one facility may have multiple requests for assessments over a period of years, with each request differing in terms of the number of classrooms to be assessed, number of teachers in each classroom, and the number of assessments needed. The data structure was designed so that additional states and/or other research partners may use the website without significant modifications.

Continuing Work (Next Steps)

The next phase of work on EQuIPS will continue the planned refinements and revisions to the measure based on data and expert review and develop a plan for a larger validation study. Before the larger validation, the revised measure will be tested in a small number of centers and public schools in North Carolina and Delaware. The larger validation study will need to include additional states to further test the psychometric properties of the measure and links to other constructs of quality within the field. Initial work on the use of the measure in public school settings began in the fall of 2016 and next steps include further testing of a revised administrator and teacher interview. The major next step activities important to continue moving the EQuiPS measure forward are outlined below:

1. Revise measure based on further analyses.
2. Pilot test the revised version of EQuIPS observation in 30 programs (including centers, homes and public schools) to refine the time sampling and global rating methodology as well as the outdoor checklist.
3. Pilot testing a revised version of the administrator and teacher interview, and the document review protocol with public school Pre-K programs, center-based programs, and homes.
4. Plan a multi-data validation study based on the revised EQuIPS measure
5. Refine the website for multi-state use and develop initial training materials for use of the website
6. Create an application process for additional states to partner with North Carolina and Delaware for a validation study of EQuIPS

References

BAS; Talan, T.N., & Bloom, P.J. (2009). *Business Administration Scale.* New York, NY: Teachers College Press.

Burchinal, M. R., Roberts, J. E., Riggins, R., Zeisel, S. A., Neebe, E., & Bryant, D. (2000). Relating quality of center-based child care to early cognitive and language development longitudinally. *Child Development, 71*(2), 339-357.

Burchinal, M., Howes, C., Pianta, R., Bryant, D., Early, D., Clifford, R., & Barbarin, O. (2006). Predicting child outcomes at the end of kindergarten from quality of pre-kindergarten teacher-child interactions and instruction. *Applied Developmental Science, 12*(3), 140-153.

Cassidy, D., Hestenes, L., Hedge, A., Hestenes, S., & Mims, S. (2005). Measurement of quality in preschool child care classrooms: The Early Childhood Environment Rating Scale-Revised and its' psychometric properties. *Early Childhood Research Quarterly, 20(3)*, 345-360.

Early, D.M., Bryant, D.M., Pianta, R.C., & Clifford, R.M. (2006). Are teachers' education, major, and credentials related to classroom quality and children's academic gains in pre-kindergarten? *Early Childhood Research Quarterly, 21*(2), 174-195.

ECERS-R; Harms, T., Clifford, R.M., & Cryer, D. (1998) *Early Childhood Environment Rating Scale-Revised Edition*. New York, NY: Teachers College Press.

FCCERS-R; Harms, T. Cryer, D., & Clifford, R.M. (2007). *Family Child Care Environment Rating Scale-Revised Edition*. New York, NY: Teachers College Press.

ITERS; Harms, T. Cryer, D., & Clifford, R.M. (2003). *Infant/Toddler Environment Rating Scale-Revised Edition*. New York, NY: Teachers College Press.

Hamre, B., La Paro, K.M., Pianta, R.C., & LoCasale-Crouch, J. (2014). *The Classroom Assessment Scoring System Infant Manual*. Baltimore, MD: Brookes Publishing.

Hestenes, L., Cassidy, D., Hegde, A., & Hansen, J. (2007). Quality in inclusive and non-inclusive infant and toddler classrooms. *Journal of Research in Childhood Education, 22*(1), 69-84.

La Paro, K.M., Pianta, R.C., & Hamre, B. (2012). *The Classroom Assessment Scoring System*

Toddler Manual. Baltimore, MD: Brookes Publishing.

Mashburn, A.J., Pianta, R.C., Hamre, B.K., Downer, J.T., Barbarin, O.A., Bryant, D.….Howes, C. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development, 79*(3), 732-749.

PAS; Talan, T.N., & Bloom, P.J. (2011). *Program Administration Scale-Second Edition*. New York, NY: Teachers College Press.

Pianta, R., La Paro, K., & Hamre, B. (2008) *The Classroom Assessment Scoring System Pre-K Manual*. Baltimore, MD: Brookes Publishing.

Schaack, D., Vi-Nhuan, L., Claude, M. S. (2013). Examining the factor structure of the Family Child Care Environment Rating Scale-Revised. *Early Childhood Research Quarterly, 28*, 936-946.

Vitello, V. (2014). Using the CLASS Measure in Family Child Care Homes. Charlottesville, VA: Teachstone.