

EXECUTIVE SUMMARY

Quality measurement serves as a foundation for child care Quality Rating and Improvement Systems (QRIS). Understanding the variation that exists in quality measurement, the different contexts in which states make decisions, and the factors that play into their decisions can help administrators identify where their state may fall along the spectrum in practice (or the direction in which they should steer), and can help researchers design approaches that take these differences into account. Such information can also aid in identifying opportunities for moving toward some common practice and research goals.

The quality components included in a QRIS define a state's framework for measuring quality and signal to providers and parents the practices that should be included in high-quality early child care and education programs. There are commonalities in the quality categories that are included across QRIS (Tout et al. 2010), demonstrating that many states and communities are using a similar foundation upon which to build their rating systems. However, the manner in which states and localities combine and aggregate these quality categories to develop QRIS ratings has many nuances, producing rating systems with important variations that can impede direct cross-QRIS comparisons and research approaches.

Recognizing the need for information on the quality measurement practices in QRIS, this in-depth study of select QRIS was launched as part of the Child Care Quality Rating Systems (QRS) Assessment project, funded by the Office of Planning, Research, and Evaluation (OPRE) within the Administration for Children and Families.¹ We examined the approaches used by states and communities to measure quality through the QRIS by focusing on three research questions:

1. What is the variation in how select QRIS define and measure quality, and what accounts for the variation in their approaches?
2. What are the specific processes used by select QRIS to measure each component of the quality rating and determine the overall rating level?
3. What is the availability of consistent and reliable data on quality ratings within select QRIS and how are the data currently being used?

To answer these questions, we selected five QRIS for the in-depth study: Miami-Dade County, Florida; Illinois; Indiana; Pennsylvania; and Tennessee. A summary of key characteristics of the five QRIS is presented in Table ES.1.

¹ To conduct the QRS Assessment project, in 2008 OPRE contracted with Mathematica Policy Research, Child Trends, and Christian and Tvedt Consulting. The goals of this project are to (1) gather and analyze existing and new information on QRIS implementation and research to inform decision making on QRIS development and refinement and (2) build the capacity for ongoing monitoring and evaluation within and across systems.

Table ES.1. Overview of QRIS Participating in the In-Depth Study of Quality Measurement

	Miami-Dade County	Illinois	Indiana	Pennsylvania	Tennessee
QRIS Name	Quality Counts	Quality Counts	Paths to Quality	Keystone STARS	Star-Quality Child Care Program
Starting Year of Statewide Implementation	2008	2007	2008	2003	2001
Number of Rating Levels	5	4	4	4	3
Structure of Rating Levels	Combination	Building blocks	Building blocks	Building blocks	Combination
Eligible Programs					
Center-based	✓	✓	✓	✓	✓
Head Start/ Early Head Start	✓	✓	✓	✓	✓
Pre-kindergarten	✓	✓	n/a	✓	✓
Licensed FCC	✓	✓	✓	✓	✓
License-exempt Homes	✓	✓			
School-aged Programs	✓	✓	✓	✓	✓
Other			Child care ministries		
Total Number of Participating Programs	430	1,030	2,040	4,420	2,749
Percentage of Participants in Each Level	Level 1: 13% Level 2: 29% Level 3: 35% Level 4: 19% Level 5: 4%	Level 1: 10% Level 2: 30% Level 3: 60% Level 4: 1% Tier 1: 52% ^a Tier 2: 22% Tier 3: 27%	Level 1: 61% Level 2: 15% Level 3: 13% Level 4: 10%	Level 1: 46% Level 2: 29% Level 3: 12% Level 4: 14%	Level 1: 2% Level 2: 19% Level 3: 61%
Total Number of Children Served	28,000 (as of July 2010)	43,465 (as of April 2011)	75,993 (as of May 2011)	168,530 (as of June 2010)	Not available

Source: Compendium of Quality Rating Systems and Evaluations (Tout et al. 2010); QRS Data (Illinois Department of Human Services, May 2011); Keystone STARS 2010 Program Report (OCDEL, 2010); Tennessee Report Card & Star Quality Program Year 8 Annual Report (Pope and Magda 2010); Paths to QUALITY, Monthly Management Report (FSSA, 2011); Trends from Miami-Dade's QRIS (ELC, 2010).

n/a = not applicable

^a License-exempt homes have a separate three-tier system in Illinois.

Quality Definitions, Thresholds, and Rating Criteria

Each of the five QRIS looked to existing systems to inform the design of their program, but they did not adopt another existing QRIS as a whole. Instead, the design of each system was influenced by characteristics of the local early education and care market, as well as existing licensing and accreditation standards.

Selecting Quality Rating Components

Although there is some overlap in quality components included in ratings, we found considerable variation in the specificity and rigor of indicators for each component.

- **Licensing.** The role of licensing depends on the perceived rigor in the licensing requirements and the maturity of the QRIS. Licensing compliance is either a complete or partial requirement at level one, or a prerequisite for participation. Only Miami-Dade does not include licensing compliance as a requirement at any level. To ensure a level playing field for all providers, four of the five QRIS require license-exempt centers to obtain a license in order to participate in their QRIS.
- **Ratio, group size, and health and safety indicators.** QRIS requirements for child-staff ratio, group size, and health and safety were influenced by the licensing requirements in each state. Two of the five QRIS include additional requirements beyond what is required for a licensed provider for child-staff ratio and group size in their QRIS standards for center-based programs in order to bring providers in line with accreditation standards by the time they reach the highest rating level.
- **Staff qualifications.** All five QRIS incorporate staff qualifications into their ratings. Education and training are the most typical indicators used but there is wide variation in how requirements are defined at each level. For example, the number and specificity of requirements vary by position type as does the percentage of staff who must meet requirements.
- **Administration and management.** Quality indicators in the area of administration and management cover two main topics—staff management (such as staff benefits, annual professional development plans for staff, and the use of differentiated salary scales based on education and experience) and program administration (such as risk and fiscal management, program evaluation, and strategic planning). Systems vary in the specificity of requirements as well as the rating level at which each is required.
- **Family partnerships and community involvement.** All five QRIS include requirements for family partnerships and community involvement; however, they typically rely on self-reported information from providers. Communication with families is a common indicator across all five systems but there is great variation in the modes specified and frequency required.
- **Environment.** Four of the five QRIS use the Environment Rating Scales (ERS; Harms et al. 1995, 2005, 2006, 2007), citing their wide use and recognition in the field. Indiana does not use the ERS scales in their entirety, but includes some items similar to ERS items in their rating tools. The four QRIS that use the ERS integrate scores into quality ratings by setting a minimum score that providers must meet to qualify for a particular rating level in building block systems or receive a number of points in combination systems. At the highest rating level, all four QRIS require a score of 5.0 or higher which aligns with the “good” range on the scale. There is greater variation among QRIS in the minimum ERS scores required at lower levels.
- **Individualization of services.** This group of components reflects the extent to which providers tailor or individualize services to meet the needs of children and families by using child assessments, provisions for special needs, developmentally appropriate curricula, and practices that respond to and recognize cultural and linguistic diversity.

With few exceptions, standards for these components are included in the ratings at higher levels across QRIS, indicating that these features are not necessarily expected of a provider demonstrating a baseline level of quality.

- **Accreditation.** Across the five QRIS, respondents perceived that accreditation represents the high end of the child care quality spectrum. Two QRIS require accreditation to reach the highest QRIS level. While accreditation is required of providers at level four in Indiana, providers must also undergo an observational assessment and demonstrate that they meet all standards of the lower levels. Other QRIS chose not to make accreditation a requirement but use alternative ways to incorporate accreditation status into ratings. In Pennsylvania, accreditation fulfills partial requirements at level four. Miami-Dade and Tennessee QRIS do not include accreditation as a requirement at any level, but award additional points to accredited providers.

Laying Out the Structure of Quality Rating Levels

- **Methods for combining indicators.** The five QRIS combine and aggregate components using either a building block approach in which a provider must meet all of the standards required at one level before moving on to the next or combination systems, wherein a provider is rated on individual components before scores are combined to obtain an overall rating. Respondents in QRIS using a building block approach noted that this structure allows for a clear and consistent representation of how each level is defined. Conversely, planners from the QRIS that employ a combination system indicated that it was necessary to provide multiple avenues to achieve a higher rating, while still prioritizing what they felt were the most important elements.
- **Number of rating levels.** Planners' and administrators' knowledge of licensing and accreditation standards served to establish a range of quality for the QRIS to cover. They designed the levels of the QRIS to help providers progress from licensing requirements (at the base) to standards that are largely equivalent to accreditation (at the top). The number of intermediate levels in each QRIS was influenced by what planners and administrators felt were reasonable expectations in terms of improvements that providers could achieve over time and supports that could be provided to help providers make that progress.
- **Terminology for levels.** Respondents in each QRIS indicated that they devoted a substantial amount of thought and discussion during QRIS planning to what ratings should be called, such as stars or levels. Four of the QRIS use the term "stars" because the term denotes a certain level of prestige and accomplishment that they want providers to associate with being a QRIS participant. Indiana's QRIS uses the term "paths" to place more emphasis on the value of the quality improvement process.

Processes for Quality Measurement

The next step in our investigation of quality measurement was an analysis of the processes that sites implement to collect information on each component and its indicators, and to determine the overall rating level.

Pre-rating Process

- All five QRIS require attendance in overview sessions in which information is provided about the goals of QRIS, the system standards, expectations of participants, and resources and supports available to help providers at each stage of the process.
- Each site has a preparation process in place to help providers learn about QRIS standards and gauge whether they are ready to undergo the rating process. During this stage, a range of supports are offered to providers including self-study materials such as workbooks and worksheets to help understand the standards and conduct self-assessments, additional training sessions, and individualized technical assistance provided through consultation with QRIS specialists.

Gathering Evidence for Individual Components

- The first step in the formal rating process is an evidence review to determine whether a provider meets requirements for individual quality components. Across the QRIS, raters are distinct staff from QRIS specialists who perform the pre-rating and supportive roles with providers.
- Rating teams vary substantially in size due to the nature and amount of work that raters are required to do in each site.
- All five QRIS require raters to have a bachelor's degree and three require that this degree be in early childhood education or a related field. None of the five QRIS has a formal protocol for training new raters or firm guidelines for initial and ongoing reliability. However, three have developed materials to improve the consistency of the evidence review process.
- Across QRIS, raters review evidence for at least 2 and as many as 10 components for each provider. Evidence is usually obtained through direct observation, director/provider interview, document review (the most common method), or a combination of the three.
- The required evidence for some components is fairly straightforward—for example, providers need only present current certificates to demonstrate licensing compliance and accreditation status. Other components can require extensive effort or documentation. For example, demonstrating staff qualifications requires access to and review of education and training documentation for numerous individual staff members.

Conducting Assessments Using Standardized Measures

- Four of the QRIS assess the quality of the environment using the ERS. Illinois also uses the Program/Business Administration Scales (Talan and Bloom, 2004, 2009) to assess a number of other components. Indiana does not assess any components using standardized measures but includes some observational indicators in their quality rating tool. (The rest of this section focuses on ERS and thus Indiana is not included.)
- The number of assessors per QRIS ranges widely from 7 assessors in Illinois to nearly 60 assessors in Tennessee. The workload for assessors is similar across sites, with assessors conducting between 8 to 12 assessments per month.

- Assessment teams include lead assessors who supervise groups of assessors, train new assessors, and serve as anchors. Anchors' ratings serve as the benchmark upon which the ratings of other assessors are compared for consistency. Lead assessors also conduct assessments themselves, albeit with a smaller caseload.
- Assessors are required to have a minimum of a bachelor's degree, and are preferred to have this degree in early childhood education. In addition to education level, two QRIS require experience in early childhood settings and Miami-Dade requires assessors to be bilingual due to the demographics of the providers and families in their locality. Miami-Dade and Pennsylvania also gauge the writing skills of assessor candidates.
- Each QRIS has built upon publisher-provided materials and guidelines to design protocols for conducting training sessions in-house. The basic parameters of the training process are similar across sites. Few of the current assessors in the five QRIS have received direct training from the authors of the ERS. In lieu of sending all assessors for training with ERS authors, three QRIS have sent at least some of their lead assessors to receive training.

Procedures for Conducting Classroom Observations

- Information on children's ages is used to determine which measures will be used for the observation. In mixed age classrooms, sites typically use the measure appropriate for the age of the majority of children in the room.
- Three sites do not inform providers of the exact date of the visit and instead give providers a window of three to four weeks during which they can expect the assessment visit to take place, and allow providers to designate blackout dates during which they cannot be observed due to scheduling conflicts.
- Assessors typically observe one-third of the classrooms for each age group served and conduct at least one assessment for each age group. In the case of multiple classrooms, the classrooms observed are selected randomly on the morning of the assessments. There are additional guidelines for selecting classrooms for observation. For example, three of the QRIS require that at least half of enrolled children are present in a particular classroom, two QRIS exclude classrooms that are staffed by a substitute teacher, and two QRIS exclude classrooms if the teacher is new.
- To calculate facility ERS scores, Illinois and Pennsylvania take the average score across all classrooms and scales administered. Tennessee also calculates an average across classrooms. However, if any individual classroom receives an ERS score below 3.0, the entire facility assumes that classroom's score. Miami-Dade produces separate averages for each scale administered (such as an ECERS average and an ITERS average).

Assigning Component and Final Ratings

- In Indiana and Miami-Dade, component ratings are automatically calculated in QRIS databases based on data entered by the rater and/or assessor for individual indicators. In Illinois, Pennsylvania, and Tennessee, raters manually calculate ratings for each component.
- Miami-Dade, Indiana, and Tennessee have the calculation of overall ratings automated in their QRIS databases; that is, based on individual component ratings, the database

automatically calculates the overall rating. In Illinois and Pennsylvania, raters determine overall ratings manually by reviewing individual component ratings.

Data Collection, Use, and Analysis to Refine Quality Measurement in QRIS

Availability of Data on Quality Measurement Ratings

- At a minimum, each QRIS database stores information on current and historical quality ratings.
- Beyond the ratings, all QRIS databases store component-level ratings for at least some quality rating components. Three QRIS databases store indicator-level data.

Use of Data to Monitor and Evaluate QRIS

- Administrators examine distributions of quality ratings at least annually to examine how QRIS participants are progressing. Administrators also examine data at the component or indicator levels to identify areas showing substantial progress and areas where large numbers of providers tend to underperform.
- Several sites monitor the supports that providers access in preparation for the rating process. Respondents discussed plans to eventually link these data to quality improvements made over time to determine which components have required the most support from QRIS Specialists.
- Of the sites we visited, only Indiana had a study already in process to compare developmental outcomes of children in the care of providers with varying quality rating levels. Respondents in other QRIS also expressed an interest in examining relationships between quality ratings and child outcomes but noted that the cost of conducting child assessments was prohibitive.
- Respondents noted that the scarcity of time and resources and the need for better integration and more detailed information on quality and outcomes are the key challenges in using the available data for research and evaluation purposes.

Research Directions

This in-depth study describes what is conceptualized as quality and how it is measured in five QRIS. Although the five QRIS profiled in this report incorporate a greater number of components in quality ratings than earlier iterations of QRIS, there remain many unanswered questions about which quality components to include, and how, within the rating systems.

In terms of quality measurement processes, we found greater consistency in the administration of the ERS across QRIS than in the procedures for gathering evidence on other quality components or calculating ratings. Nonetheless, there continue to be threats to the reliability of standardized assessments including limitations in the number of assessors trained directly by authors of the measures and inconsistencies in the number of classrooms observed. The measures of other quality components present challenges to consistent, reliable data collection and interpretation. Multiple modes of data collection—such as observation, interview, and document review—could serve to confirm the presence of quality components (and increase reliability) but would likely introduce tradeoffs in terms of cost.

Among the five QRIS studied, there is generally greater consistency in the definitions of the quality components at the highest rating levels than at the baseline levels. Cut-off points at intermediate levels are somewhat arbitrarily determined. Whether differences between providers at each level would translate to meaningful differences in child outcomes is an open question. At the highest level, QRIS standards overlap considerably with recommendations of accrediting organizations such as the National Association for the Education of Young Children. Further research may help shed light on whether features specified for the highest level are consistent with quality thresholds that have been linked to positive outcomes for children.