



Chapter Title: Introduction

Book Title: Assessing the Validity of the Qualistar Early Learning Quality Rating and Improvement System as a Tool for Improving Child-Care Quality

Book Author(s): Gail L. Zellman, Michal Perlman, Vi-Nhuan Le and Claude Messan Setodji

Published by: RAND Corporation

Stable URL: <https://www.jstor.org/stable/10.7249/mg650qel.9>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



This content is licensed under a RAND Corporation License. To view a copy of this license, visit <https://www.rand.org/pubs/permissions.html>.



JSTOR

RAND Corporation is collaborating with JSTOR to digitize, preserve and extend access to Assessing the Validity of the Qualistar Early Learning Quality Rating and Improvement System as a Tool for Improving Child-Care Quality

The Effects of Quality

Research findings in recent years point to the importance of the preschool period in children's longer-term development. These findings have focused attention on the quality of care young children are receiving outside their homes, a focus reinforced by the growing numbers of young children cared for by non-family members (Lamb, 1998; Scarr, 1998; Vandell and Wolfe, 2000). These concerns about quality have been abetted by a policy focus in the K–12 sector on students' academic achievement. Growing concerns about performance outcomes in elementary school have led policymakers and others to examine the degree to which early education promotes school readiness and improves children's longer-term academic performance.

Studies of the effect of child care on children's outcomes have focused on the quality of that care. Child-care quality is generally viewed as encompassing both structural and process characteristics. Structural characteristics include staff training and education, child-staff ratios, and aspects of the physical environment. Process elements involve the quality of child-staff interactions and instruction. Structural characteristics tend to be more quantifiable and, therefore, more amenable to regulation than process characteristics, which are harder to quantify and, therefore, regulate. Despite the greater challenges in measuring and regulating it, process quality is considered more critical, as it influences children more directly. Structural characteristics are viewed as driving the quality of the processes that take place in a given setting.

Numerous studies have demonstrated that higher-quality child care is predictive of a range of positive developmental outcomes for children, including improved language development, cognitive functioning, social competence, and emotional adjustment (e.g., Howes, 1988; National Institute of Child Health and Human Development Early Child Care Research Network [NICHD ECCRN], 2000; Peisner-Feinberg et al., 2001; Burchinal et al., 1996; Clarke-Stewart et al., 2002). However, the magnitude of these effects has begun to be debated. The strongest effect sizes are reported for studies in which disadvantaged children are randomly assigned to programs that provide high doses of high-quality care and extensive supports to parents in a very prescribed way (Ramey and Ramey, 2006). These children are then compared with those who were randomly assigned to the "no organized care" condition. For example, the Carolina Abecedarian Project (Campbell and Ramey, 1995) and the High/Scope Perry Preschool Project (Weikart, Bond, and McNeil, 1978) report effect sizes for I.Q. greater than .60 over time compared with no care. This long-term effect of child-care quality on developmental outcomes for disadvantaged children is generally agreed to reflect the fact that high-quality child-care programs provide learning opportunities and social and emotional support that may not be available at home (Heckman, 2006; Scarr, 1998). Detailed studies of parent-child inter-

actions in families of different income levels reinforce this notion (e.g., Hart and Risley, 1995). But because of their designs, these studies do not speak to the effect of gradations in the quality of care children receive and the effects of these gradations on children's outcomes.

Consistent with the above argument, studies with more demographically varied samples report more moderate effect sizes.¹ The strongest evidence from nonexperimental studies suggests that the effect sizes in studies that examine the relationship between child-care quality and child functioning are fairly small. For example, in a study by the National Institute of Child Health and Human Development (NICHD) and Duncan (2003) notable for its rigorous methods, with children between 2 and 4.5 years of age, the effect sizes were between .04 and .08. They conclude that "child care quality is a modest but reliable predictor of cognitive development and academic achievement during early childhood" (p. 1470). A recent study of pre-academic achievement in state pre-kindergarten (Pre-K) programs found that enrollment in these programs was associated with statistically significant gains in some academic and social skills, but the gains were small. Moreover, most classrooms lacked the process-quality components associated with such gains (Howes et al., 2008).

Some studies have found no link at all between child-care quality and child outcomes. For example, Deater-Deckard, Pinkerton, and Scarr (1996) failed to find a relationship between the quality of preschool child care and school-age children's social, emotional, or behavioral adjustment. Scarr (1998) suggested that family effects, confounded with child-care quality, account for long-term results observed in other longitudinal studies. A Dutch retrospective study (Goossens, Ottenhoff, and Koops, 1991) also reported no effects of child-care quality on development and achievement in school-age children. A more thorough Swedish study (Broberg, Hwang, and Chace, 1993) reported similar findings. This latter study was conducted in a country with "uniformly high-quality child-care centers," and therefore does not provide the range in provider quality that would enable a fair assessment of the relationship between quality and child outcomes, and does "not really test for the effects of poor child care on later development" (Scarr, 1998, p. 104).

All of these findings must be considered against a strong bias in the literature toward publishing only significant results. Little work has examined this bias; none of it speaks directly to the effect of child-care quality on children. However, Roggman et al. (1994) conducted a search of unpublished studies on child care and mother-child attachment and reported that many of these studies found null results (i.e., no relationships between child care and mother-child attachment). The authors conclude that assumptions about the effect of child care on attachment would be weaker if these unpublished data were considered. Analogously, it is reasonable to assume that findings about the effect of child-care quality on child functioning would be weaker if a similar search were conducted on the effects of child-care quality on child outcomes.

Despite these mixed findings about the effects of child-care quality on child outcomes, there continues to be widespread consensus that quality matters. Some argue that quality care changes children's trajectories. We argue that quality care is good for young children on a day-to-day basis whether or not it is associated with long-term improvements in their cognitive or social functioning. Rich learning environments, supportive interactions with adults, and scaf-

¹ Furthermore, most studies involve nonrandomized designs; self-selection bias and differential attrition may also be influencing effect sizes (Ramey and Ramey, 2006).

folding that encourages exploration are all good things for children, regardless of whether they affect their subsequent outcomes.

The importance of quality and its relationship to children's daily experiences and longer-term outcomes takes on added urgency in light of consistent research findings that much child care is mediocre at best (e.g., Peisner-Feinberg and Burchinal, 1997; NACCRRA, 2006b). Quality and cost are closely related. Lower child-staff ratios and better-educated staff are generally viewed as two key elements of quality (e.g., NACCRRA, 2006b; Zellman and Gates, 2002). Both are major cost drivers. For example, lower child-staff ratios for younger children raise the cost of infant care. Zellman and Gates (2002) found that the cost of providing infant care in accredited Department of Defense child development centers was almost twice that of providing high-quality care to a preschooler in the same center. While every state provides some child-care assistance, which partially subsidizes the cost of care for some families, many working families are not eligible, and others who are eligible face long waiting lists (NACCRRA, 2006b).² Given low child-care subsidies and the inability of parents to absorb fee increases, children most at risk in terms of school readiness are likely to be found in lower-quality care.³

The generally low quality of child care has led to calls to improve quality, amid recognition that the current child-care system in the United States, if it can be called a system at all, does little to promote it. While much care is licensed, licensing represents a fairly low quality bar, since it focuses on the adequacy and safety of the physical environment. The limited amount of care in many locations and for key age groups (particularly infants) generally provides ready clients for most providers, whether or not they offer quality care. This strong demand for spaces at any quality level limits provider incentives to take often-costly steps to improve. In some cases, providers may not know how to improve, even if they are motivated to do so. There are few empirical data available that providers can call on to help them select the best ways to invest limited quality-improvement (QI) funds in order to maximize increases in quality. Another constraint on QI may be found in parents' limitations in recognizing high-quality care and distinguishing it from care of moderate or mediocre quality. Although some believe that quality is obvious and that parents will "know it when they see it," research described below suggests that parents may not know what to look for, and, even if they do, they may make care decisions based on other, more pressing considerations. Some argue that parents may mistakenly use fees as an indicator of quality because they do not know how to make an independent assessment (Zellman and Perlman, forthcoming). It may also be possible that parents do not value the same "quality" characteristics that researchers value. This hypothesis is advanced by Kisker and Maynard (1991), who note that provider education and training, ratios, and curricula may seem less important to parents than the provider's personal characteristics, such as warmth or the newness and brightness of the facility.

² The National Association of Child Care Resource and Referral Agencies (NACCRRA, 2006b) reports substantial gaps between the earnings limit to receive child-care assistance and the earnings necessary to purchase average-priced child care in the four least affordable states.

³ A significant exception to the association between cost and quality may be found at Head Start centers and at Child Development Centers sponsored by the Department of Defense for military dependents. In both of these settings, substantial subsidies enable low-income children to receive care of high quality at very low cost (Zellman and Gates, 2002; U.S. Department of Health and Human Services, 2004).

Efforts to Improve Quality Through Quality Rating Systems

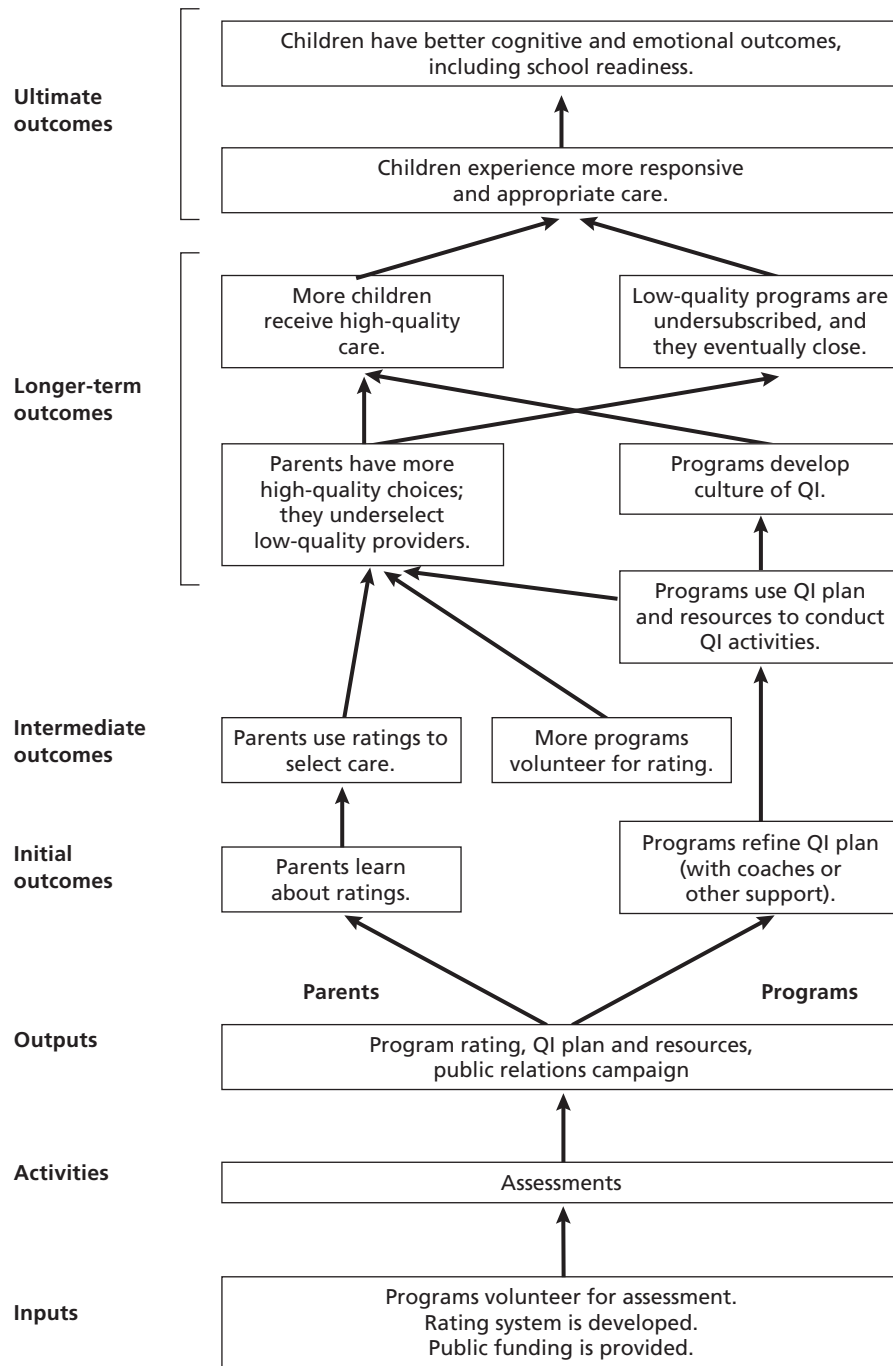
Although the challenges are daunting, there have been a number of attempts to intervene in the child-care system to promote quality improvements. One approach that has been embraced widely of late involves the development and implementation of quality rating systems (QRSs). (See Zellman and Perlman, forthcoming, for further description of these systems and the importance of having adequate resources.) QRSs are assessments based on multiple qualitative indicators designed to make child-care quality transparent and easily understood. Most of these QRSs are really QRISs—quality rating and *improvement* systems. QRISs have been promoted because it became apparent early on that, on the provider side of the equation, motivation alone would not improve quality. Many child-care providers lack the ability to determine how to improve; the summary ratings that are the outputs of the QRS assessment process help little on their own. Providers need more-detailed assessments and a quality-improvement plan. Moreover, quality improvements cost money. In particular, lower child-staff ratios and better-trained staff, two components that are generally viewed as critical to quality, are major cost drivers.

QRISs, therefore, provide hands-on technical assistance and QI resources to participating providers to improve the level of quality they offer. This hands-on technical assistance is closely linked to the results of the multicomponent QRIS assessment; these systems often produce a detailed QI report, in addition to a summary rating. Such support creates incentives for providers to be assessed and rewards providers for doing better in an accountability system that promotes quality improvement and more-informed parental choices. QRISs generally adhere to the logic model shown in Figure 1.1.

The general theory underlying QRISs is that child-care quality is difficult to ascertain. Therefore, QRISs focus on creating an assessment system that produces a single, easy-to-understand rating for each provider. These ratings make quality transparent for parents, providers, funders, and other stakeholders. This is important because both economic theory and research argue that if there is adequate supply and parental ability to pay fees, parent choices that are based at least in part on quality could drive quality of care (e.g., Gormley and Weimer, 1999). Once these assessments are available, the logic model posits that parents will use them to select the highest-quality care that they can afford, and providers will be motivated to improve their quality so that they can improve their rating. Such motivation is theorized to arise from the desire to run a prestigious program and to fill all available spaces in the program. Support for quality improvement, e.g., staff training, often contributes to provider motivation. For programs that receive subsidies, tiered reimbursement—a payment system in which staff and providers rated as having higher quality receive higher per-child subsidies—may provide additional incentive to improve. Higher-quality programs, a long-term outcome, are posited to enhance the everyday environment for children. An improved environment, characterized by more-responsive caregiving and enriched content, will then lead to better outcomes for children. These outcomes may include improved school readiness, cognitive skills, and non-cognitive outcomes, such as social skills development and creativity.

While the specifics of each system vary, in general, participating providers are assessed on each of the system components (typically 4–7) and receive a summary rating that they are encouraged to display. These simple, readily understood ratings (often 0–5 stars or a rating of 1–4), are posited to increase the ability of parents, funders, and other stakeholders to make

Figure 1.1
A Logic Model for QRISs



RAND MG650-1.1

more-informed choices about which providers to use and support.⁴ When parents can easily ascertain quality, the theory is that they will choose more effectively and efficiently, selecting the highest level of quality that they can afford; those parents who are less financially constrained will be able to choose care of the highest quality. The motivating force of the QRIS theoretically extends to providers as well. Faced with the prospect of putting a placard with just 1 star (when 4 are possible) in the front window of their center, QRIS theory posits that participating providers will be motivated to improve the quality of the care they offer.⁵

The idea behind QRISs is compelling, although there are significant concerns about whether such systems can actually work, given the realities of U.S. child care. In many locations, there is not enough care to meet the need, particularly if one needs specific kinds of care, e.g., infant care (NACCRRRA, 2006a). Lack of supply limits the effect of demands for improved quality. Chipty (1995) found that many child-care providers meet, but do not exceed, state licensing standards. One reason that they fail to exceed minimal standards, he contends, is that higher standards increase the cost of care. When costs increase, providers have two unattractive options: absorbing the additional cost or raising the price of care. When prices increase, parents generally purchase less care. Either way, Chipty argues, providers do not benefit financially from providing higher-quality care. In other locations, high-quality care may be available but is very costly; parents may prefer it but cannot afford to purchase it.

Parents may not choose care of the highest quality even if they can afford it because other factors may dominate the decisionmaking process. One such factor is convenience. For nearly all families, someone must make two visits each day to the provider—to drop off and pick up the child. If the care is located far from work or home, location can become a challenge (Gates et al., 2006). Hours of operation also can create problems for parents. For example, some centers do not open early enough in the morning to accommodate parents whose jobs begin very early (Zellman and Johansen, 1996). Some research (e.g., Johansen, Leibowitz, and Waite, 1997) finds that location and price are the key characteristics that parents report they consider in choosing child care.

Assessing and Validating QRISs

The theory underlying QRISs has yet to be tested. Indeed, there is little information available about these systems—how well they measure what they purport to measure, whether parents pay attention to ratings in selecting care, whether providers that participate in QRISs actually improve the quality of the care they provide, and whether children benefit from the improved care they are receiving as their provider receives quality-improvement support.

Many of the existing systems are based on consensual ideas about what components of quality are most important in creating a program that supports child development. The component measures themselves have been assessed infrequently, and their combination into summary measures and particularly the manner in which they are weighted to determine a summary quality rating have little empirical basis. Studies of K–12 high-stakes accountability

⁴ Morris and Helburn (2000) found that suppliers sometimes supply lower quality at the same price as higher-quality services and can get away with it because of parent ignorance. (See also Helburn and Bergmann, 2002.)

⁵ Given the voluntary nature of virtually all QRISs, it is reasonable to assume that those that do participate are motivated to improve, believe that they already provide high-quality care, or both.

systems clearly show that when test scores have stakes attached to them, test-takers attend to what is on the test (Corbett and Wilson, 1991; Shepard and Dougherty, 1991). It is likely that child-care providers will respond in the same way when they are rated in a high-stakes context characterized by public ratings and consequences associated with those ratings. Therefore, it is critical that the right constructs be captured in these QRISs. Principles of fairness to child-care providers, parents, funders, and other QRIS users make it imperative that the components of a QRIS measure what they purport to measure. It is also worthwhile to determine whether there are less labor-intensive ways to assess quality. If assessment costs could be reduced, more money might be available for quality improvement.

A number of the QRIS systems in place have conducted evaluations of selected parts of their systems. However, these evaluations have focused, for the most part, on a single issue: whether summary ratings are correlated with a single widely used measure of quality, the Early Childhood Environment Rating Scale–Revised (ECERS-R; Harms, Cryer, and Clifford, 1998) and its component measures. (See Zellman and Perlman, forthcoming, for further discussion of evaluations of QRISs in five early-adopting states.) But there have been no systematic efforts to validate QRIS components or the summary ratings that constitute the major outputs of QRISs.

What does it mean to validate a quality rating system? *Validity* refers to the degree to which evidence and theory support the conclusions derived from multicomponent assessments. The validation process is necessarily quite specific: Validation must be focused on a specific purpose and is limited to a specific context. With QRISs, the purpose is to assess how well the system components measure child-care quality in a context in which considerable stakes may be attached to a particular rating. It is important to note that validity is not attached to just a measure, but to a measure used for a particular purpose or in a particular context. This means that measures that may be valid for one use must be validated again for use in a different context (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 1999). This is a particular issue with QRISs, because at least some of the most widely used measures of child-care quality, e.g., the ECERS-R, were developed and have been used in contexts with low stakes. Measures developed in low-stakes contexts must be validated again in high-stakes contexts because providers being assessed in high-stakes contexts may react in ways that may undermine the meaningfulness of interpretations derived from those measures.⁶

Validation is a complex, iterative process. A thorough validation process requires that multiple sources of evidence be collected. These may include expert judgments concerning the degree to which measurement instruments capture the components of quality and whether individual items are consistent with the domain being assessed. They may also include quantitative data. The pattern of relationships among the scores on different measures of the same concept (including the one being investigated) and the pattern of relationships among the items within a measure are two of the most important types of validity evidence that can be collected. For example, measures of a given quality domain, such as child-staff interactions, should relate more closely to each other than to measures of other quality domains, such as the physical environment (Campbell and Fiske, 1959).

⁶ For example, in high-stakes contexts, those being assessed may focus improvement efforts on the most easily modified aspects of a measure (e.g., number of books in the ECERS-R) while ignoring other aspects.

Scores on a given measure may also be compared with other variables. For example, since higher levels of staff education and training are believed to result in more child-centered interactions, scores on these measures should be related.

The purpose of this study was to investigate the validity of a major initiative designed to improve child-care quality in Colorado, the Qualistar QRIS, which we will refer to hereafter as the Q-QRIS to differentiate it from generic QRISs. This effort, which includes variables representing all the categories in the QRIS logic model, represents the first empirical investigation of a QRIS.

Qualistar's QRIS

Qualistar Early Learning, a Colorado nonprofit, designed a QRIS, which was first implemented in 1999.⁷ The Q-QRIS, developed to assess child-care provider quality, was intended to help improve quality by including indicators that would generate useful feedback for quality improvement. This meant that the measures had to be sufficiently detailed and “actionable” enough to form the basis of a quality-improvement plan. Quality is conceptualized as a multidimensional concept represented by the five Q-QRIS components, displayed in Table 1.1. Several of these components are themselves multidimensional measures, e.g., the classroom environment rating and parent involvement (e.g., family partnerships) measure. The Q-QRIS is designed to assess quality in classrooms serving infants, toddlers, and preschoolers, as well as family child-care homes.

The rating system components were derived from the Cost Quality and Outcomes Study (Peisner-Feinberg et al., 1999; Peisner-Feinberg and Burchinal, 1997; Helburn et al., 1995), which had a major effect on Colorado child-care policy. It also relied on its own founders' extensive experience and sense of what mattered in creating high-quality care. There is considerable consensus in the field that most of these components contribute to high-quality care. Child-staff ratios; the size of the group in which a child receives care; staff education, training, and experience; characteristics of the physical environment; and caregiver-child interactions have all been viewed as very important (National Research Council and Institute of Medicine, 2000b). Since Qualistar's goal was to create a high-stakes system that would eventually be used to determine levels of public funds flowing to centers, it understood that the components that were included would be those that people paid attention to. This led system developers to include parent involvement, at the time a not-well-developed concept, because Qualistar believed that parents should have some role in their child's provider.

The Q-QRIS includes the five components shown in Table 1.1, and relies on a point system in which providers must earn specified numbers of points to qualify for a particular star rating.⁸ The maximum total score is 42 points, distributed equally across components (10 points per component), with the exception of accreditation, which is worth 2 points; the program receives no points on this component if it is not accredited. A program's point totals are

⁷ Qualistar Early Learning, formerly Educare, is a Colorado nonprofit supported by the Colorado Trust, the Temple Hoyne Buell Foundation, the Boettcher Foundation, the Daniels Fund, the Rose Foundation, the Denver Foundation, and the Chambers Family Fund.

⁸ Point systems allow providers to focus their QI efforts on areas where they think they can maximize points (see Zellman and Perlman [forthcoming] for more discussion of point systems and the other frequently used approach, block systems).

Table 1.1
Qualistar Early Learning QRIS Components

| Component | Description | Points Possible |
|--|---|-----------------|
| Classroom learning environment ^a | Measured using the ECERS-R and the Infant/Toddler Environment Rating Scale (ITERS). ^a The ECERS-R point total is 7 for each of the seven subscales: space and furnishings personal care routines language-reasoning activities interaction program structure parents and staff (averaged across classrooms). | 10 |
| Child-staff ratios ^b | Number of children and staff in each room at the time of assessment. Measured in a variety of ways over the course of this study. Averaged across classrooms. | 10 |
| Staff and director training and education (T&E) ^a | Based on the following three pieces of information for each staff member: years of experience in child care formal education Early Childhood Education (ECE) credits. These are averaged across staff in settings that have multiple staff: 7 points for teachers (averaged across teachers). 3 points for directors (averaged across directors). | 10 |
| Family partnerships ^c | Measured in a variety of ways over the course of the study. Generally includes a score based on parent surveys and another score based on documentation and other information from center director. Collected for parents in infant/toddler as well as preschool; points earned based on % of parents who scored activities as occurring; points earned per item based on evidence of activity occurring. | 10 |
| Accreditation ^c | Identifies whether or not the provider has been accredited by a national accrediting agency. | 2 |

NOTE: Qualistar Early Learning permits accreditation by other agencies besides the National Association for the Education of Young Children (NAEYC), but most accredited programs work with the NAEYC system; this was true of virtually all providers in this sample.

^a ITERS was not included in the evaluation but was included in provider QRIS score.

^b Collected at the classroom level.

^c Measured at the center level.

combined to produce a summary star rating of 0–4 stars, as shown in Table 1.2. In addition to a summary star rating, providers received a detailed overall profile of their program and a quality-improvement plan based on the assessment. Qualistar coaches work with providers to develop and refine the QI plan and implement it in their program. (See Document A.1.1 in the appendix to this chapter for further description of Qualistar’s QRIS.)

Many sources of information feed into any validation effort, including review of past literature, expert opinion, and empirical data, as discussed previously. Our validation effort included all of these information sources. Our effort focused on analyzing the relationship among the Q-QRIS component measures and the relationship of the Q-QRIS components to measures of quality external to the Q-QRIS. We assessed the relationship between Q-QRIS star ratings, the ultimate system outcome, and child outcomes.

Table 1.2
Qualistar Early Learning Star Levels Criteria

| Star Rating: Provisional | Centers: 0–9 Points or Learning Environment Score of 0 | Homes: 0–9 Points or Learning Environment Score of 0 |
|-----------------------------|--|--|
| Star 1 | 10–17 points | 10–16 points |
| Star 2 | 18–25 points | 17–23 points |
| Star 3 | 26–33 points | 24–30 points |
| Star 4 | 34–42 points | 31–39 points |

Research Questions

The remainder of the report presents our efforts to examine the components of the Q-QRIS and validate the ratings as an indicator of child-care quality and as a tool to improve. These analyses were driven by the following research questions:

1. What are the characteristics of the Q-QRIS components as measures?
2. How closely related are the five Q-QRIS component measures?
3. Do providers that receive high scores on the Q-QRIS components also receive high scores on process-quality measures (the Caregiver Interaction Scale [Arnett, 1989] and the Pre-Kindergarten Snapshot [Howes, 1997]) that were chosen as criteria?
4. Is there a relationship between the star ratings and the individual Q-QRIS components and concurrent child outcomes? Is provider quality related to future child outcomes?
5. Which Q-QRIS components contribute most to child outcomes?
6. How should the components be combined into a Q-QRIS in a way that takes into account the relative contributions of the components to child outcomes?
7. Are there subgroups of children for whom the links between measures of child-care quality and child outcomes are stronger?
8. Did child-care quality improve over time?

Organization of This Report

In Chapter Two, we present study methods. Chapter Three presents analyses and discussion of the five Qualistar QRIS components, respectively. In Chapter Four, we present data on the relationships among the Q-QRIS components and present an analysis of quality improvement over the course of the study in participating providers. In Chapter Five, we present the results of the models that link the rating system components and the summary star rating to other measures of quality and to child outcomes. In Chapter Six, we examine the family child-care providers included in the study in some detail. Because of the small numbers of homes in the study sample and the small numbers of eligible children in each home, we could not conduct the more complex analyses that we applied to center data. For this reason, we chose to analyze the child-care home data separately.

In Chapter Seven, we discuss our findings and contextualize them through a literature review of comparable studies. We draw implications from this work for the Qualistar QRIS and for the development of quality rating and improvement systems by others. In particular,

we discuss what the field needs to do to create better QRISs as one means of improving child-care quality.

Finally, the CD that accompanies this report includes seven appendixes, each corresponding to a chapter in the report. These appendixes contain supporting documents and additional data, as explained in each chapter.

